

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|-----|--------|----------|----|--------------------|
| | | Match | | | | | |
| 1 | 1429 | 100.0 | 548 | 21 | AAV44633 | | Human organic cati |
| 2 | 342.5 | 24.0 | 439 | 22 | AAV76766 | | Corynebacterium gl |
| 3 | 342.5 | 24.0 | 448 | 22 | AAV93041 | | C glutamicum prote |
| 4 | 338 | 23.7 | 742 | 21 | AAV94977 | | Human secreted pro |
| 5 | 283.5 | 19.8 | 542 | 22 | AA947274 | | HOAT3. Homo sapie |
| 6 | 279.5 | 19.6 | 542 | 21 | AAV92902 | | Human cerebell org |
| 7 | 278 | 19.5 | 536 | 21 | AAV92903 | | Rat cerebell organ |
| 8 | 278 | 19.5 | 556 | 17 | AAV77676 | | Rat OCT-1 protein. |
| 9 | 271.5 | 19.0 | 535 | 21 | AAV51249 | | Rat liver anion tr |
| 10 | 270.5 | 18.9 | 537 | 18 | AAW44195 | | Mouse osteoclast t |
| 11 | 268.5 | 18.8 | 561 | 18 | AAW44196 | | Human osteoclast t |

Db 34 ssgwaldamdvgllsifvmaalathwgsptetsilgsifvgmaigsaglladklgr 93
 QY 80 KTGGLAISVLWTLYGI---LSAFAPVYSWILVRLGLVGGGG-VPOSVTLVYAEFLPWKA 135
 Db 94 r---qvafsllyvvgatgaalsvslamlnalrfvvgiglaelpvastlissetsprkv 150
 QY 136 RAKCILLIEVFAIGTVEVVLAVFVMPSL--GWRWLLIILSAVPLLLFAVLFCFWLPESAR 193
 Db 151 rgrmvvileatwalgwlnaaivgtftvagsdngwrwalagcvpaivayvrglpesvr 210
 QY 194 YDVLSSNGEKA---LATLKRRTATENGAPMPGLKLIISROEDRGKMRDLFTPHFRWTTLL 250
 Db 211 flekgrhdeaealvvsfeaaagkaadattavvhdnaegsv-siwsaalrkrvtal 269
 QY 251 WPIFNSAFSYG 263
 Db 270 wivwfcinslyyg 282

RESULT 4

AAAY94977
 ID AAY94977 standard; Protein: 742 AA.

XX AC AAY94977;

XX DT 16-JUN-2000 (first entry)

XX DE Human secreted protein clone as180_1 protein sequence SEQ ID NO:160.

XX KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antihypertoid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy.

XX OS Homo sapiens.

XX PN WO200009552-A1.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18298.

XX PR 14-AUG-1998; 98US-0096622.

XX PR 17-AUG-1998; 98US-0096815.

XX PR 04-SEP-1998; 98US-0099229.

XX PR 23-OCT-1998; 98US-0105368.

XX PR 08-JAN-1999; 99US-0115234.

XX PR 12-FEB-1999; 99US-0119931.

XX PR 18-FEB-1999; 99US-0120575.

XX PR 30-APR-1999; 99US-0132020.

XX PR 11-AUG-1999; 99US-0096622.

XX PA (GEM) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

XX PI Wong GG, Clark HF, Fechtel K;

XX DR WPI; 2000-205979/18.

XX PT New polynucleotides encoding secreted proteins, which may have e.g.

XX PT nutritional, chemokine, immune stimulating or suppressing.

XX PT hematopoiesis regulating, tissue growth, activin/inhibin

XX PT antiinflammatory or tumor inhibition activity

XX PS Claim 169; Page 614-616; 641pp; English.

XX CC AAL16618 to AAL16697 encode the human secreted proteins given in
 CC AAY94988 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAL16698 to AAL16774 represent
 CC probes for the human secreted proteins from the present invention.

XX SQ Sequence 742 AA;

Query Match 23.7%; Score 338; DB 21; Length 742;

Best Local Similarity 27.4%; Pred. No. 1.7e-29;

Matches 86; Conservative 54; Mismatches 126; Indels 48; Gaps 5;

QY 1 EDAYEALGFCFKFQWLKSLVTGLAWMADAMWMLLSILAPQLHCEWRLPWSQVALLTSVVF 60

Db 153 ealirecghrfqwtlyfvgialmadgvefvvvgfvlpsaekdmcldsdnkgmglivy 212

QY 61 VGMSSSTLNGNISDQYGRKTGLKISVLWLYGILSAFAPVYSWILVRLGLVGGIGG- 119

Db 213 lgmvmgafllwggialrqrqcllisvnsvfaffsfvgyggtfllcrllsgvgiggs 272

QY 120 VPOSVTLVYAEFLPMKARAKCILLIEVFAIGTVEVVLAVFVMPSLG----- 166

Db 273 ipivfsyfeqlaqekrghlswlcmfmwlggyaaamawailphygwsfqmgsayqfhs 332

QY 167 WRWLLLSAVPLLLFAVLFCFWLPESARYDVLISGNOEKAIATLKRIATEN---CAP---M 219

Db 333 wrvflvcafpsvfaigalttqpesprfllengkhdeawmvkqvhdtnmrakghpervf 392

QY 220 PLGKLIISROEDR-----GKMRDLFTPHFRWTTLLMF 252

Db 393 svthikthgedelieqlsdgtgygrvraislsggvwgnflscfgpeyrritlmmmg 452

QY 253 IFFSNAFSYGVLV 266

Db 453 wvftmsfsyygltv 466

RESULT 5

AAB47274

ID AAB47274 standard; Protein: 542 AA.

XX AC AAB47274;

XX DT 06-AUG-2001 (first entry)

XX DE hOAT3.

XX KW Human; organic anion transporter; hOAT; liver; kidney;

XX KW membrane protein; transport; organic anion; splice variant.

XX OS Homo sapiens.

| YX | Sequence | 556 AA; |
|-----|----------|---------|
| 1 | ... | ... |
| 2 | ... | ... |
| 3 | ... | ... |
| 4 | ... | ... |
| 5 | ... | ... |
| 6 | ... | ... |
| 7 | ... | ... |
| 8 | ... | ... |
| 9 | ... | ... |
| 10 | ... | ... |
| 11 | ... | ... |
| 12 | ... | ... |
| 13 | ... | ... |
| 14 | ... | ... |
| 15 | ... | ... |
| 16 | ... | ... |
| 17 | ... | ... |
| 18 | ... | ... |
| 19 | ... | ... |
| 20 | ... | ... |
| 21 | ... | ... |
| 22 | ... | ... |
| 23 | ... | ... |
| 24 | ... | ... |
| 25 | ... | ... |
| 26 | ... | ... |
| 27 | ... | ... |
| 28 | ... | ... |
| 29 | ... | ... |
| 30 | ... | ... |
| 31 | ... | ... |
| 32 | ... | ... |
| 33 | ... | ... |
| 34 | ... | ... |
| 35 | ... | ... |
| 36 | ... | ... |
| 37 | ... | ... |
| 38 | ... | ... |
| 39 | ... | ... |
| 40 | ... | ... |
| 41 | ... | ... |
| 42 | ... | ... |
| 43 | ... | ... |
| 44 | ... | ... |
| 45 | ... | ... |
| 46 | ... | ... |
| 47 | ... | ... |
| 48 | ... | ... |
| 49 | ... | ... |
| 50 | ... | ... |
| 51 | ... | ... |
| 52 | ... | ... |
| 53 | ... | ... |
| 54 | ... | ... |
| 55 | ... | ... |
| 56 | ... | ... |
| 57 | ... | ... |
| 58 | ... | ... |
| 59 | ... | ... |
| 60 | ... | ... |
| 61 | ... | ... |
| 62 | ... | ... |
| 63 | ... | ... |
| 64 | ... | ... |
| 65 | ... | ... |
| 66 | ... | ... |
| 67 | ... | ... |
| 68 | ... | ... |
| 69 | ... | ... |
| 70 | ... | ... |
| 71 | ... | ... |
| 72 | ... | ... |
| 73 | ... | ... |
| 74 | ... | ... |
| 75 | ... | ... |
| 76 | ... | ... |
| 77 | ... | ... |
| 78 | ... | ... |
| 79 | ... | ... |
| 80 | ... | ... |
| 81 | ... | ... |
| 82 | ... | ... |
| 83 | ... | ... |
| 84 | ... | ... |
| 85 | ... | ... |
| 86 | ... | ... |
| 87 | ... | ... |
| 88 | ... | ... |
| 89 | ... | ... |
| 90 | ... | ... |
| 91 | ... | ... |
| 92 | ... | ... |
| 93 | ... | ... |
| 94 | ... | ... |
| 95 | ... | ... |
| 96 | ... | ... |
| 97 | ... | ... |
| 98 | ... | ... |
| 99 | ... | ... |
| 100 | ... | ... |

Query Match 19.5%; Score 278; DB 17; Length 556;
Best Local Similarity 32.0%; Pred. No. 7.5e-23;
Matches 72: Conservative 48; Mismatches 95; Indels 1

| | | | |
|----|-----|---|-----|
| QY | 49 | SNQVALLTSVVVGMMSSSTLWGNISDQYGRKTKGLKISVLTLYVYGLISAFAPVSWILV | 108 |
| Db | 146 | awkvdifgscvnlgfvlgsivvgviadrfrkicllvtltvtsvsgvltavapdysmll | 205 |
| QY | 109 | LRGLVGF-GIGVQSQVTLVAEFLPMKARAKACILLIEVFWAIGTVFEVVLAVFMPSLQW | 167 |
| Db | 206 | frllqgmwsgswsgvyllitefvsgsyrirtaillyqmaftvgvlagva-yaipd--w | 262 |
| QY | 168 | RWLLILSAVPLLLFAVLCLFVPLPSARYDVLSGNQEKAIATLKRIATENGACWPLGLKLIIIS | 227 |
| Db | 263 | rwlqlavslplrfilyyvwfvpesprwlisqgktrtravimeidqkndkvvpadlkmic | 322 |
| QY | 228 | RQEDRGKMR-----DLF-TPHERWTTLLLWTFWFSNAFSYGYGLVL | 266 |
| Db | 323 | leedasekrspsfadlfrtlrkrhtvilmvylwfscavlvgqlim | 367 |

RESULT 9
AAY51249
ID AAY51249 standard; Protein; 535 AA.

DT 14-APR-2000 (first entry)

Rat liver anion transporter protein OAT2.

KW OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaqlandin;
KW non-steroid anti-inflammatory agent; anti-tumor.

OS *Rattus* sp.

PN JP11346779-A.

PD 21-DEC-1999.

PF 03-JUN-1998; 98JP-0169174.

03-JUN-1998; 98JP-0169174.

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN

WPI; 2000-109694/10.

DR N-PSDB; AAZ44679.

PT A liver-specific org. anion transporter - can transport various
PT substances including dicarboxylic acids
PT

PS Claim 2; Page 5-6; 13pp; Japanese.

CC This invention describes a novel rat liver-specific org. anion
CC transporter OAT2. The liver-specific org. anion transporter OAT2
CC can transport various substances including dicarboxylic acids,
CC prostaglandins, non-steroid anti-inflammatory agents and anti-tumor
CC agents. This sequence represents the rat liver anion transporter OAT2
CC protein described in the method of the invention.

| Sequence | 535 AA; |
|----------|---------|
| SQ | |

Query Match 19.0%; Score 271.5; DB 21; Length 535;
Best Local Similarity 30.6%; Pred. No. 3.9e-22;
Matches 81; Conservation 47; Mismatch 4

QY 16 LSVLTGLAWNADAMENMILSILAPQLHCEWRLPSWQVAL--LTSV-VFVGMSSSSTLWGN 72

| | | | |
|----|-----|---|-----|
| Db | 109 | ltvpcsgwgydrse-----fstlataewdlvcqgrginkltstcfftgvlvgavvyg | 162 |
| Qy | 73 | ISDQYGRKTKLISVLTWLYGILSAPFYSHILVRLGLVGIGGVQSQV-TLXAEFL | 131 |
| Db | 163 | lsdfgrrrlllvayvssylvlgimsaasinymfvvtrltgtsalagftlvipllewl | 222 |
| Qy | 132 | PMKARAKILLIEVFEWAIGVFVEVLAVFVMPISLGRWHTLLISAVPLLLFVLCFWLPES | 191 |
| Db | 223 | dvehrtvagvstvfws-gvlllvalgyliirs--wrwlllaatpcvpqgilsiwvwpes | 279 |
| Qy | 192 | ARDVVLSGNQEKAITLKRIATENGAMPPLGKL-----IISROEDRGKMRDLF-TP | 241 |
| Db | 280 | arwltqtgrveeakkyllscakngirpvgegslsqealnvvumeralqrpysldlrts | 339 |
| Qy | 242 | HFRWTTLLLLWFIFWFSNAFSYIGLVL | 266 |
| Db | 340 | qlrhisiccmwvfwqvnfsyyqtl | 364 |

RESULT 10

AAW44195
ID AAW44195 standard; Protein; 537 AA.

AA AAW44195;
AC

DT 14-MAY-1998 (first entry)

XX Mouse osteoclast transporter protein. DE

XX
KW
Mouse; osteoc

XX
OS
Mus musculus.

XX PN W09742321-A1

XX PD 13-NOV-1997

XX
PF 09-MAY-1997.XX
PR 09-MAY-1996.XX
PA (RGHM) BRIGXX
PT
Reier DP
DXX
WP1: 1997-55

DR N-PSDB; AAV1:
YY

PT Novel human a
isolating

PT
yy

PS Disclosure; R
yy

The present study

CC protein can be

CC and osteopetr

| | | |
|----|----------|----|
| SQ | Sequence | 53 |
|----|----------|----|

| | | | | |
|-----------------------|-------|--------------------|--------|-------------|
| Query Match | 18.9% | Score 270.5; | DB 18; | Length 537; |
| Best Local Similarity | 24.6% | Pred. No. 5.1e-22; | | |

| | | | |
|----|----|---|----|
| Qy | 2 | DAVEAIG-EGKQKLSVLTGLANMADAME--MMILSILAPQLHCE-----WRLP- | 48 |
| | | : :: : : : : : : | |
| Db | 5 | elldrvsgmgpfqylhvtllalpilgltannlllqftattpdhcrrpppnaslepwvlp | 64 |
| Qy | 49 | -----SVOVA-----L | 54 |

Db 65 gngkpkcklrfvhlpnaslpndtqgatepclgdgwiynstrdtivtewdlvcgsnkikem 124
 QY 55 LTVSVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAFAPVYSWILVRLGLVG 114
 Db 125 aqsfmagilvgpvfgeclsdrgkrpiltwsllylaasgssafspstlvymifrlcog 184
 QY 115 FGIGGVPOS-VTLYAEFLPMKARAKCILLIEVFMAICTVFEVVLAVFVMPISLG-----WR 168
 Db 185 csisglstlilnvwvptsttraiststtligcyctigq-----filpglayavpqr 236
 QY 169 WLLILSAVPLLLFAVLCFWLPESARYDVLGSGNOEKAATLKRAT-----ENGAPMLGK 223
 Db 237 wlqisvsaafifsliswwpesirwlvjsgkfsralktlqrvtatfngkkeegekitvee 296
 QY 224 LIISROED-----RGKMRDLF-TPHFRWTTLLWFWFNSAFSYYGLVL 266
 Db 297 lknldqkdisakvkgysldfrvsilrrvtfclsawfatgfayyslam 346

RESULT 11

AAW44196
 ID AAW44196 standard; Protein; 561 AA.

XX AC AAW44196;

XX DT 14-MAY-1998 (first entry)

XX XX Human osteoclast transporter protein.

XX DE Human; osteoclast transporter protein; osteoporosis; osteopetrosis.

XX KW Homo sapiens.

XX OS WO9742321-A1.

XX PN 13-NOV-1997.

XX PD 09-MAY-1997; 97WO-US07856.

XX PF 09-MAY-1996; 96US-0647397.

XX PR (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX PI Beier DR, Brady KP;

XX XX WPI; 1997-558983/51.

DR N-PSDB; AAV12393.

XX PT Novel human and mouse osteoclast transporter proteins - useful for
 PT isolating agents that decrease osteoclast activity, for treatment of
 PT osteoporosis and osteopetrosis

PS Disclosure: Page 35-39; 52pp; English.

XX CC The present sequence represents human osteoclast transporter protein.
 CC Agents that bind the nucleic acid encoding the osteoclast transporter
 CC protein can be used to decrease its expression, thereby decreasing
 CC osteoclast activity. The agents can be used to treat osteoporosis
 CC and osteopetrosis.

XX SQ Sequence 561 AA;

Query Match

Best Local Similarity 18.8%; Score 268.5; DB 18; Length 561;

Matches 86; Conservative 49; Mismatches 122; Indels 93; Gaps 12;

QY 2 DAVEAIG-FGRFOWKLSVLTGLAW--MADAMEMMLSLILAPQLHCE-----WRLP- 48

Db 5 eildrvsgmhgfihvailygilnmanhnlqiftaatpvhcrrphnastgwpvlpm 64

QY 49 -----SQOVA-----L 54

Db 65 gngkpkcklrfvhlpnaslpndtqgatepclgdgwiynstkdsvtewdlvcnsnkikem 124
 QY 55 LTVSVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAFAPVYSWILVRLGLVG 114
 Db 125 aqsfmagilvgilvgldisdrfgrpiltwsllylaasgssafspstlvymifrlcog 184
 QY 115 FGIGGVPOS-VTLYAEFLPMKARAKCILLIEVFMAICTVFEVVLAVFVMPISLG-----WR 168
 Db 185 fgisgltstlilnvwvptsttraiststtligcyctigq-----filpglayavpqr 236
 QY 169 WLLILSAVPLLLFAVLCFWLPESARYDVLGSGNOEKAATLKRAT-----ENGAPMLGK 222
 Db 237 wlqitvsiptfifflsswwtpeirwlvjsgksskalkilrrvgclqwgqgrrrkislee 296
 QY 223 -KLIIISROEDRGKMR-----DLF-TPHFRWTTLLWFWFNSAFSYYGLVL 266
 Db 297 lknldqkdisakvkgysldfrvsilrrvtfclsawfatgfayyslam 346

RESULT 12

AAW49401

ID AAB49401 standard; Protein; 540 AA.

XX AC AAB49401;

XX DT 02-MAR-2001 (first entry)

XX DE Murine organic anion transporter 6.

XX KW Murine; organic anion transporter 6; mOATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder; kidney disease;
 KW liver disease; autoimmune disease.

XX OS Mus sp.

XX PN WO200070048-A1.

XX PD 23-NOV-2000.

XX PF 15-MAY-2000; 2000WO-US13316.

XX PR 14-MAY-1999; 99US-0134137.

PR 12-MAY-2000; 2000US-0570293.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Feild J, Yue L, Ellens H;

XX WPI; 2001-016235/02.

DR N-PSDB; AAC83979.

XX PT Murine organic anion transporter 6 polypeptide, useful for identifying
 PT agonists/antagonists that are useful in treatment of cancer, kidney
 PT disease, autoimmune disease, inflammation and cardiovascular disease -
 XX Claim 2; Page 28-29; 32pp; English.

XX CC The present sequence is murine organic anion transporter 6 (mOATP6).
 CC mOATP6 protein is useful for screening compounds which inhibit or
 CC stimulate the function of mOATP6 and also compounds that neither agonise
 CC nor antagonise OATP6. The identified agonists and antagonists are useful
 CC for prevention and treatment of human diseases, including cancer,
 CC inflammation, cardiovascular disease, central nervous system disorders,
 CC kidney diseases, liver disease and autoimmune diseases.

XX SQ Sequence 540 AA;

Query Match

Best Local Similarity 17.9%; Score 256.5; DB 22; Length 540;

Matches 76; Conservative 47; Mismatches 117; Indels 23; Gaps 8;

```
QY 16 LSVLTCLAWADAMEMMILSLAPOLHCEWRLPSSQVAL---LTSVVFVGMSSSTLWGN 72
Db 109 ltvpcsqweydrse-----fstiatewdlvcdqrglnkvtstccffigvilgavvvey 162
QY 73 ISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGGVQSV-TLYAEFL 131
Db 163 lsdrgfrrllvaystlaiglmsaasvnyimfvttmllgsalagftiilvplewl 222
QY 132 PKKARAKCILLIEVFAIGTVFEVVLAVFVMPSLQWRWLLLSAVPLLLFAVLCFWLPES 191
Db 223 dvehrtvqvisttwt-ggvllltlvgyllrs--wrwlllaatlpcvpgiiswwvps 279
QY 192 ARYDVLGSGNQERKAIATLKRATPENGAPM-----PLCKLI-ISRQEDRGKMRDLF-TP 241
Db 280 arwlitqgrveaekylsacalnrgpisedsalsqeaInkvitmervsqrpsyldlrts 339
QY 242 HFRWTTLLWFWFNSAFSYGVLV 266
Db 340 qirhvsllccmmwfwgnfsyygltl 364

RESULT 13
AAW64538
ID AAW64538 standard; Protein: 554 AA.
XX
AC AAW64538;
XX
DT 21-OCT-1998 (first entry)
XX
DE Human liver cell clone HP01293 protein.
XX
KW Transmembrane domain; human: nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
PN WO9821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX
PR 13-NOV-1996; 96JP-0301429.
XX
PA (PROT-) PROTEGENE INC.
PA (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR WPI: 1998-297932/26.
DR N-PSDB; AAV49558, AAV49559.
XX
PT Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
PS Claim 1; Page 96-98; 205pp; English.
XX
CC AAW64534-W64558 represent human proteins containing a transmembrane
CC domain. These proteins can be used for, e.g. research and nutrition, and
CC may have cytokine and cell proliferation/differentiation, immune
CC stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokine, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
SQ Sequence 554 AA;
```

Query Match 17.8%; Score 255; DB 19; Length 554;
Best Local Similarity 31.1%; Pred. No. 3, 1e-20;
Matches 70; Conservative 42; Mismatches 103; Indels 10; Gaps 5;

```
QY 49 SNOVALLTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILV 108
Db 145 swkldfscinagffgslvgvyfadrfgrklcllgtvlvnavsgvlmafspnymsmll 204
QY 109 LRLGVGF-GIGVQSVSTLYAEFLPKKARAKCILLIEVFAIGTVFEVVLAVFVMPSLCW 167
Db 205 frllqglvskgnmagtyltitefvgsgsrtrvaimyqmaftvglvaltgla-yalph--w 261
QY 168 RWLLTILSAVPLLLFAVLCFWLPESARYDVLGSGNQERKAIATLKRATPENGAPMPLCKLIIS 227
Db 262 rwlqavslptflfilywvcpesprwlisqkrnteakimdhiaqknkplpadikmls 321
QY 228 ROED-----RGKMRDLF-TPHFRWTTLLWFWFNSAFSYGVLV 266
Db 322 leedvteklspsfadlfrtprlkrftfilmylwfldsvlyqglil 366

RESULT 14
AAB76717
ID AAB76717 standard; Protein: 399 AA.
XX
AC AAB76717;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:416.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-Q141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
```

```

PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042080.
XX
XX
PA (BADI ) BASF AG.
XX
XX Pompejus M, Kroege B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-071486/08.
XX N-PSDB; AAF67950.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
XX and membrane transport proteins or their portions, useful for typing or
XX identifying C. glutamicum or related bacteria, and as markers for
XX transformation -
XX
XX Claim 20; Page 759-760; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
XX construction and membrane transport (MCT) proteins given in AAB/6510 to
XX AAB/6847. The MCT nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria (e.g. Brevibacterium lactofermentum), the typing or
XX identification of C. glutamicum or related bacteria, as reference points
XX for mapping C. glutamicum genome, and as markers for transformation.
XX AAF68082 and AAF68082 represent sequencing primers which are used in an
XX example from the present invention.
XX
XX Sequence 399 AA;
XX
XX Query Match 17.2%; Score 246.5; DB 22; Length 399;
XX Best Local Similarity 27.4%; Pred. No. 1.8e-19;
XX Matches 71; Conservative 47; Mismatches 126; Indels 15; Gaps 7;
XX
XX Qy 16 LSVLTGLAWMA---DAMEMMILSILAPOL--HCEWRLPSMOVALLTSVFVGMSSSTLW 70
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 7 vgtvlalwfaivdvgfdvlgvlgatpsmledpawdltagatqistglvgmtigalti 66
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Qy 71 GNISDQYGRKTKGLKISVLWTLYYGILSAFAPVYSWILVRLGLVGFGIIGG-VPOSVTLVAAE 129
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 67 gftldrlgrrrvmfsvavsvftlllafttnvqlfslwrlflagvlggalptaiaamvte 126
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Qy 130 FLPMKARAKCILLIEVFNAICTVEVVLAVFVMPSLGHRWLLIILSAVP-LLLFAVLCPFWL 188
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 127 frpgtkagsattlmtgyhvgavataflglidgfgwhsmfigagvpgllilpllyffl 186
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Qy 189 PESARYDVLNSNOEKATLKRRIATENGAPMLPKLIISROEDRGK---MRDLFTPHFRW 245
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 187 pespaylkisgkldaeqa---vaasyglsid-dlddreheeelegssslsfkpsfir 241
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Qy 246 TLLLLWFIWFNSAFSYGL 264
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 242 ntlaiwgtsgmglilvygl 260
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 15
XX AAG90943
XX ID AAG90943 standard; Protein; 431 AA.
XX
XX AC AAG90943;
XX
XX XX
XX DT 26-SEP-2001 (first entry)
XX
XX DE C glutamicum protein fragment SEQ ID NO: 4697.
XX
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX OS Corynebacterium glutamicum.
XX

```

| | |
|----|---|
| PN | EP1108790-A2. |
| XX | |
| PD | 20-JUN-2001. |
| XX | |
| PF | |
| XX | 18-DEC-2000; 2000EP-0127688. |
| XX | |
| PR | 16-DEC-1999; 99JP-0377484. |
| PR | 07-APR-2000; 2000JP-0159162. |
| XX | 03-AUG-2000; 2000JP-0280988. |
| XX | (KYOW) KYOWA HAKKO KOGYO KK. |
| PA | |
| XX | Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H; |
| PI | Tateishi N, Senoh A, Ikeda M, Ozaki A; |
| XX | |
| DR | WPI; 2001-376931/40. |
| DR | N-PSDB; RAAH66162. |
| XX | |
| PT | Novel polynucleotides derived from Coryneform bacteria, for identifying |
| PT | mutation point of a gene, measuring expression of a gene, analysing |
| PT | expression profile or pattern of a gene and identifying homologous gene |
| XX | |
| XX | Claim 17; SEQ ID NO: 4597; 245pp + Sequence Listing; English. |
| XX | |
| CC | The present invention provides a number of nucleotide and protein |
| CC | sequences from the Coryneform bacterium <i>Corynebacterium glutamicum</i> . These |
| CC | are useful for identifying the mutation point of a gene derived from a |
| CC | mutant of coryneform bacterium, measuring expression amount and |
| CC | analysing the expression profile or expression pattern of a gene derived |
| CC | from coryneform bacterium, and identifying a homologue of a gene deriving |
| CC | from coryneform bacterium. Coryneform bacteria are useful for producing |
| CC | amino acids, nucleic acids, vitamins, saccharides and organic acids, |
| CC | particularly L-lysine. The present sequence is a protein described |
| CC | in the exemplification of the invention. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification, but was obtained in electronic format directly from the |
| XX | European Patent Office. |
| XX | |
| SQ | Sequence 431 AA; |
| | |
| | Query Match 17.2%; Score 246.5; DB 22; Length 431; |
| | Best Local Similarity 27.4%; Pred. No. 2e-19; |
| | Matches 71; Conservative 47; Mismatches 126; Indels 15; Gaps 7 |
| Qy | 16 LSVLITGLAWMA--DAMEMMILSILAPOL--HCEWRLPDSOVALTTSVVVFGMMSSSTLW 70 |
| Dd | LSVTLLLVFAIVDGFIVIGATIPMSLMEDPAWDLTAGYATQISTIGLVGMTIGAITI 72 |
| Qy | 71 GNISDYQGRKTKLSIVLWTYYGILSAFAPYYSWLVRGLGVGGIGG-VPOSVTLXAE 129 |
| Dd | GILTRLRVRRVIMFSVAVSFVTLILAFTINNVQLFSIWRLFIAGVGLGGALTAIAMTE 132 |
| Qy | 130 FLPHKARAKCILLIEVFVAICTVFVEVLAVFVMPSLGNRWLLILISAVP-LLLFAVLCLFWL 188 |
| Dd | FRTGTAKGSASTTMTGYHVGAATAFLIGLIDFGVHSMFIAGAVPGVLIPLIPLYIFLI 192 |
| Qy | 189 PESARYDLVSNGQEKAIAATLKRIATENGAPMLPKGLIKLISROEDRGK---MRDLTFPHFRW 245 |
| Dd | PESAPYIKISKLDLEAQ-----VAASYGLSID-ddldreheelgeesslsflfksfir 247 |
| Qy | 246 TTLLLWFTHFSNAFSYYGL 264 |
| Dd | NTLAIWGTSFMQLLVYGL 266 |

Search completed: March 13, 2002, 12:41:07
Job time: 54 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:43:20 ; Search time 21.65 Seconds
(without alignments)
283.760 Million cell updates/sec

Title: US-09-911-667A-4
Perfect score: 1429
Sequence: 1 EDVAEAGFGKFWKLSVLT.....WFSNAPSYYGLVLLTTELQ 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 281 | 19.7 | 556 | 3 | US-08-501-572-1 |
| 2 | 281 | 19.7 | 556 | 3 | US-09-040-444-1 |
| 3 | 271 | 19.0 | 555 | 3 | US-08-501-572-3 |
| 4 | 271 | 19.0 | 555 | 3 | US-09-040-444-3 |
| 5 | 270.5 | 18.9 | 537 | 2 | US-08-647-397-2 |
| 6 | 255 | 17.8 | 553 | 3 | US-08-501-572-2 |
| 7 | 255 | 17.8 | 553 | 3 | US-09-040-444-2 |
| 8 | 154.5 | 10.8 | 584 | 2 | US-08-928-692-13 |
| 9 | 145.5 | 10.2 | 524 | 2 | US-08-928-692-12 |
| 10 | 138 | 9.7 | 493 | 4 | US-09-031-392-10 |
| 11 | 138 | 9.7 | 493 | 4 | US-09-299-549-10 |
| 12 | 138 | 9.7 | 500 | 4 | US-09-031-392-7 |
| 13 | 138 | 9.7 | 500 | 4 | US-09-299-549-7 |
| 14 | 131.5 | 9.2 | 509 | 2 | US-09-031-392-6 |
| 15 | 131.5 | 9.2 | 509 | 4 | US-09-299-549-6 |
| 16 | 130.5 | 9.1 | 534 | 4 | US-09-031-392-4 |
| 17 | 130.5 | 9.1 | 534 | 4 | US-09-299-549-4 |
| 18 | 129.5 | 9.1 | 520 | 4 | US-08-964-127-2 |
| 19 | 128 | 9.0 | 492 | 2 | US-08-355-844-3 |
| 20 | 128 | 9.0 | 492 | 5 | PCT-US93-05704-13 |
| 21 | 124 | 8.7 | 286 | 4 | US-08-964-127-4 |
| 22 | 118.5 | 8.3 | 488 | 2 | US-08-928-692-11 |
| 23 | 106 | 7.4 | 494 | 2 | US-09-031-392-5 |
| 24 | 106 | 7.4 | 494 | 4 | US-09-299-549-5 |
| 25 | 105 | 7.3 | 387 | 2 | US-08-872-302-5 |
| 26 | 104.5 | 7.3 | 396 | 2 | US-08-850-880-4 |
| 27 | 104.5 | 7.3 | 396 | 2 | US-08-944-916-4 |

| | | | | | | |
|----|-------|-----|-----|---|-------------------|--------------------|
| 28 | 104.5 | 7.3 | 396 | 2 | US-08-814-877-4 | Sequence 4, Appli |
| 29 | 104.5 | 7.3 | 457 | 2 | US-08-882-704A-6 | Sequence 6, Appli |
| 30 | 103.5 | 7.2 | 455 | 1 | US-08-035-928-2 | Sequence 2, Appli |
| 31 | 103 | 7.2 | 488 | 2 | US-08-928-692-10 | Sequence 10, Appli |
| 32 | 98 | 6.9 | 383 | 2 | US-09-031-392-3 | Sequence 3, Appli |
| 33 | 98 | 6.9 | 383 | 4 | US-09-299-549-3 | Sequence 3, Appli |
| 34 | 96 | 6.7 | 514 | 1 | US-08-063-552-13 | Sequence 13, Appli |
| 35 | 96 | 6.7 | 514 | 5 | PCT-US93-05704-13 | Sequence 13, Appli |
| 36 | 95.5 | 6.7 | 563 | 2 | US-09-031-392-2 | Sequence 2, Appli |
| 37 | 95.5 | 6.7 | 563 | 4 | US-09-299-549-2 | Sequence 2, Appli |
| 38 | 94 | 6.6 | 521 | 1 | US-08-063-552-2 | Sequence 2, Appli |
| 39 | 94 | 6.6 | 521 | 5 | PCT-US93-05704-2 | Sequence 2, Appli |
| 40 | 93 | 6.5 | 515 | 1 | US-08-063-552-4 | Sequence 4, Appli |
| 41 | 93 | 6.5 | 515 | 5 | PCT-US93-05704-4 | Sequence 4, Appli |
| 42 | 93 | 6.5 | 528 | 2 | US-08-403-852D-21 | Sequence 21, Appli |
| 43 | 93 | 6.5 | 528 | 3 | US-08-510-646B-22 | Sequence 22, Appli |
| 44 | 93 | 6.5 | 528 | 4 | US-09-231-818-21 | Sequence 21, Appli |
| 45 | 90.5 | 6.3 | 423 | 2 | US-08-494-907-14 | Sequence 14, Appli |

ALIGNMENTS

RESULT 1
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics andNor Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-501-572-1

Query Match 19.7%; Score 281; DB 3; Length 556;
Best Local Similarity 32.0%; Pred. No. 1.8e-22;
Matches 72; Conservative 49; Mismatches 94; Indels 10; Gaps 5;
QY 49 SQVALLTSVVFVGMSSSTLWGNISDOYGRKTKLSVLTWLYYGLSAPFVYSWILY 108

Db 146 ANKVDLFQSCVNLGFFGLSVVGYIADRGKLCLLVTTLVTSVSGVLTAFAVDYTSMLL 205
QY 109 LRGLVGF-GIGVPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVVLAVFVMPSLGW 167
Db 206 FRLLQGWVSKGWSVGYTLITEFVGSYRRRTTALYQMAFTVGLVGLAGVA-YAIPD--W 262
QY 168 RWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIATLARIATENGAPMPLGKLIIS 227
Db 263 RWLQVAVSLPTFLFLLYYWFVPESPRWLLSQKRTTRAVRIMEQIAQKNGKVPADLKMIC 322
QY 228 ROEDRGKMR-----DLF-TPHFRWTTLLLWFIWFSNAFSYYGLVL 266
Db 323 LEEDASEKRSPSFADLFRTPNLKRHTVILMYLWFSCAVLYQGLIM 367

RESULT 2
US-09-040-444-1
; Sequence 1, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-1

Query Match 19.7%; Score 281; DB 3; Length 556;
Best Local Similarity 32.0%; Pred. NO. 1.8e-22;
Matches 72; Conservative 49; Mismatches 94; Indels 10; Gaps 5;

QY 49 SMOVALLTSVVFVGMSSSTLWGNISDOYGRKTKGLKISLVLTLYYLSAFAPVSWILY 108
Db 146 ANKVDLFQSCVNLGFFGLSVVGYIADRGKLCLLVTTLVTSVSGVLTAFAVDYTSMLL 205
QY 109 LRGLVGF-GIGVPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVVLAVFVMPSLGW 167
Db 206 FRLLQGWVSKGWSVGYTLITEFVGSYRRRTTALYQMAFTVGLVGLAGVA-YAIPD--W 262

QY 168 RWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIATLARIATENGAPMPLGKLIIS 227
Db 263 RWLQVAVSLPTFLFLLYYWFVPESPRWLLSQKRTTRAVRIMEQIAQKNGKVPADLKMIC 322
QY 228 ROEDRGKMR-----DLF-TPHFRWTTLLLWFIWFSNAFSYYGLVL 266
Db 323 LEEDASEKRSPSFADLFRTPNLKRHTVILMYLWFSCAVLYQGLIM 367
RESULT 3
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-3

Query Match 19.0%; Score 271; DB 3; Length 555;
Best Local Similarity 31.9%; Pred. NO. 2.2e-21;
Matches 74; Conservative 43; Mismatches 91; Indels 24; Gaps 8;

QY 49 SMOVALLTSVVFVGMSSSTLWGNISDOYGRKTKGLKISLVLTLYYLSAFAPVSWILY 108
Db 146 SMLDLFQSSVNVGFFIGNSIGYIADRGKLCLLVTLNAAAGVLMASISPTTWMLI 205
QY 109 LR---GLV---GFGIGVPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVVLAVFV 160
Db 206 FRLLQGLVSKAGWLIGY-----LITEFVGGRYRRTVGIFYQVAYTVGL---LVLAGVAY 257
QY 161 VMPSLGWRLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIATLARIATENGAPM 220
Db 258 ALPH--WRWLQFTVALPNFFFLYYWCIPESPRWLISQNKNAEMRIKHIKNGKSLP 315
QY 221 LCKLII:HQEDRGK-----MRDLF-TPHFRWTTLLLWFIWFSNAFSYYGLVL 266
Db 316 ASLQRLRLEETGKGLNPSFLDLVRTPQIRKHTMILMYNFTSSVLYQGLIM 367

APPLICANT: Beiter, David R.
APPLICANT: Brady, Kevin P.
TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/647,397
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gates, Edward R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: B0801/7048
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-647-397-2

Query Match 18.9%; Score 270.5; DB 2; Length 537;
Best Local Similarity 24.6%; Pred. No. 2.4e-21;
Matches 86; Conservative 51; Mismatches 120; Indels 93; Gaps 11;

Qy 2 DAVEAIG-FGKQKLSVLTGLAMADAME--MMILSLAPOLICE-----WRLP- 48
Db 5 EILDVSGMGPQYLHVHTLLALPILGIANHLLQIFTATPDHHCRRPPPNASLEPMVLP 64
Qy 49 -----SNOVA-----L 54
Db 65 GPNKPEKCLRFVHLPNLSLNDTOGATEPCLDGHIYNSTRDTIVTEWLVGSKNKLKEM 124
Qy 55 LTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVGLVG 114
Db 125 AQSVFMAGILVGGPVFGEISDRFGRKPILTWSYLLLAASGSAFSPSLTYMIFRFLCG 184
Qy 115 FGIGVPOS-VTLXAEFLPMKARAKILLIEVFVAIGTVFVVLAVFVMPSLG-----WR 168
Db 185 CSISGISLTIILNVWVPTSTRAISSTTGYCYTIGO-----FILPLAYAVPQWR 236
Qy 169 WLLILSAVPLLLFAVLCFWLPESARYDVLGNGQEKATLKRAT-----ENGAPMPLGK 223
Db 237 WLQLSVSAAFFIFLSLWVWPESIRWLVLGSKFSRAKTLQRTVATFNGKKEGKLTVEE 296
Qy 224 LIISQED-----RGKMRDLF-TPHFRWTTLLLWIFWFSNAFYSYGLVL 266
Db 237 LKFNLOKQDITSAKVYGLSDLFVRVILRRVTFCLSLAWFATGATFYISLAM 346

RESULT 6
US-08-501-572-2
; Sequence 2, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin

APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein Which Effects The
TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,444
FILING DATE: March 18, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-040-444-3

Query Match 19.0%; Score 271; DB 3; Length 555;
Best Local Similarity 31.9%; Pred. No. 2.2e-21;
Matches 74; Conservative 43; Mismatches 91; Indels 24; Gaps 8;

Qy 49 SNOVALLTSVFPVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILV 108
Db 146 SNNLDFOSSVNVGFFIGMSIGYIADRFGRKLCLLTTLVNLNAAAGVLMASPTYTWMLI 205
Qy 109 LR---GLV---GFGIGVPOSVTLXAEFLPMKARAKILLIEVFVAIGTVFVVLAA--VF 160
Db 206 FRLOGLYSKAGWLICYI-----LITEFVGGRYRTVGIFQVAYTVGL--LVLGWAY 257
Qy 161 VMSLGRWLLTSLAVPLLLFAVLCFWLPESARYDVLGNGQEKATLKRATENGAPMP 220
Db 258 ALPH--WRWLOFTVALPNFFLLYYWCIPESPRWLIISONKNAEAMRIKTKIAKNGKSLP 315
Qy 221 LKGLIISROEDRGK-----MRDLF-TPHFRWTTLLLWIFWFSNAFYSYGLVL 266
Db 316 ASLQRLLEETGKGLNLSFLDLVTPQIRKHTMLMYNWTSSVLYOGLIM 367

RESULT 5
US-08-647-397-2
; Sequence 2, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:

[illegible]

```

US-09-299-549-10

Query Match          9.7%; Score 138; DB 4; Length 493;
Best Local Similarity 25.0%; Pred. No. 6.7e-07;
Matches 59; Conservative 43; Mismatches 88; Indels 46; Gaps 9;

QY   4  VEAIGFKFOWKLSVLTGLAWADAMEMMILSIAPOLHCWEHLPS-----WQVALLT 56
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   20  IAAALGSFOYGNYLGV-----NAPQKVTEAFVETWLGRGXGXPSTPTLLTW---SLSV 70

QY   57 SVVFVGHWSSSTLWGNISDOYGRKTGLKISVLMTLYYGILSAFAP---YSWILVLRGI 113
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   71 SIFAVGGMGISFLVGXIGNRGLRKXAMLVNVIATAGGLMLGAAXSEPMILJRGFI 130

QY   114 GFGTGGVPQSWTLY-AEFLPMKARAKCILLIEFWAIGTVFE--VVLAVVMPMSIG--- 166
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   131 GLYCGLSSGVVPMYVGVEISTALRG-----ALGTINQLGIVIGIIAQVLGLDSL 180

QY   167 -----WRWLIILISAVPLLLFAVLFCWLPESARYDVLSGNQE-----KAITLKRIA 212
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   181 LGNESIWPPLLIGTGVPAILOLLLPCPCSPRYLLINKNEERAKKALORLRTA 236

```

```

RESULT 12
US-09-031-392-7
: Sequence 7, Application US/09031392
: Patent No. 5942398
: GENERAL INFORMATION:
: APPLICANT: Tartaglia, Louis A.
: APPLICANT: Weng, Xun
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES
: TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Meiklejohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 500 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-031-392-7

```

Query Match 9.7%; Score 138; DB 2; Length 500;
Best Local Similarity 26.6%; Pred. No. 6.8e-07;
Matches 64; Conservative 35; Mismatches 102; Indels 40; Gaps 9;
QY 50 WQVALLTVVFCVGMSSSTLWGISDQYGRKTKGLKLSVLWTLYXGIL---SAFAPVYSWI 106
db 71 WSTV---SWPPFGGFGTSLVGLPLVKKPKRKGALIPNNIFSPVAILMCSSRVATSFELI 128

[illegible]

```

RESULT 14
US-09-031-392-6
; Sequence 6, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELE: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 509 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-392-6

```

| | | | | |
|-----------------------|--------|--------------------|----------------|--------------------|
| Query Match | 9.2% | Score 131.5; | DB 2; | Length 509; |
| Best Local Similarity | 28.8%; | Pred. No. 3.6e-06; | | |
| Matches | 61; | Conservative 27; | Mismatches 77; | Indels 47; Gaps 8; |

| | | | | |
|-----|-----|--|---|-----|
| QY | 53 | ALLTSVVFVCGMSSSTLWGNISDOYGRKTKLKI---- | SVLWTLTYGYLTSAFAPYYSWILV | 108 |
| | | :: | | ::: |
| Ddb | 82 | ALSVALFSVGGMTSSFVIGIISQWLGKRRKRLVNNVLAVLGSSLMGLANA--AASYEMLIL | 140 | |
| | | :: | | ::: |
| QY | 109 | LRGLVFGFGIGGVQPSVTLY--AEFLPMKARAKCILLIEVFWAIGTV--FEVVLAVFVMPSL | 165 | |
| | | :: | | ::: |
| Ddb | 141 | GRFLICAGSGLTSGLVPYVGEIAPTHLRG----- | ALGTNLQLAIVIGILIAQVL | 190 |
| | | :: | | ::: |
| QY | 166 | G----- | WRWILILSAPVLLLFVLCFWLPESARYDVLVSGNOF--KAIA TLKRIA-- | 212 |
| | | :: | | ::: |
| Ddb | 191 | GLSLLGTASLWPLLLGLTVLPALLOVLPLPCFSPSPRYLIIQNLEGPARKSLKRLTGW | 250 | |
| | | :: | | ::: |

```
Qy 213 -----TENGAMPLGKLIISR 228
      | | | | |
Db 251 ADVSGVLAELKDEKRRLERERPLSLQLLGSR 282
```

```

US-09-299-549-6
/ Sequence 6, Application US/09299549
/ Patent No. 6136547
/
/ GENERAL INVENTION:
/
/ APPLICANT: Tartaglia, Louis A.
/
/ APPLICANT: Weng, Xun
/
/ TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
/
/ TITLE OF INVENTION: GLUTEX AND USES THEREOF
/
/ NUMBER OF SEQUENCES: 10
/
/ CORRESPONDENCE ADDRESS:
/
/ ADDRESSEE: Fish & Richardson P.C.
/
/ STREET: 225 Franklin Street
/
/ CITY: Boston
/
/ STATE: MA
/
/ COUNTRY: USA
/
/ ZIP: 02110-2804
/

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

PRIOR APPLICATION DATA: 09/031,392
 APPLICATION NUMBER: 28-FEB-1998
 FILING DATE: 28-FEB-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Melkielejohn, Ph.D., Anita L.
 REGISTRATION NUMBER: 35,283
 REFERENCE/DOCKET NUMBER: 07334/072002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154

```

; SUBEX: 200134
; INFORMATION FOR SEQ ID NO: 6:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 509 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-09-299-549-6

```

| | | | | |
|-----------------------|--------|--|----------------|--------------------|
| Query Match | 9.2% | Score 131.5; | DB 4; | Length 509; |
| Best Local Similarity | 28.8%; | Pred. No. 3.6e-06; | | |
| Matches | 61; | Conservative 27; | Mismatches 77; | Indels 47; Caps 8; |
| QY | 53 | ALLTSVVVFVCGMSSSTLWGNISDQYGRKTKLKI-----SVLWLTYYGILSAFAPVYSILV | 108 | |
| Db | 82 | ALLSVAIFSVCGMSSFLIGIISOWLGRKRAMLVNNVLAVLGSSLMGLANA--AASYEMIIL | 140 | |
| QY | 109 | LRGLVFGIGGVQSVTVLY--AEFLPMKARAKCILLIEVFWAIGTV--FEVVLAVFVFWPSL | 165 | |
| Db | 141 | GRFLIGAYSGLTSGLPMVYGEIAPHRLG-----ALGTNLQLAIVIGILIAQVL | 190 | |
| QY | 166 | G-----WRWLLILSAVPLLLFAVLFCFWMLPESARYDVLGSGNOE-KAIAITLKRTA-- | 212 | |
| Db | 191 | GLESLLGATSLWPLLLAGLAVLPALLOIVLLPFCPESPRYLIYITQNLGGPARKSLKRITGW | 250 | |
| QY | 213 | -----TENGAPMPLGKLIISR | 228 | |
| Db | 251 | ADVSGVLAELKDEKKRLERERPLSLQLQIGSR | 282 | |

Search completed: March 13, 2002, 12:43:21

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:41:38 ; Search time 25.88 Seconds
(without alignments)
803.541 Million cell updates/sec

Title: US-09-911-667A-4
Perfect score: 1429
Sequence: 1 EDVAEAIQFGKQWKLKSLVT.....WFSNAPSYGLVLLTTELFQ 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----------|--------------------|
| 1 | 831.5 | 58.2 | 529 | 2 T23190 | hypothetical prote |
| 2 | 707 | 49.5 | 300 | 2 S15786 | glucose transport |
| 3 | 446.5 | 31.2 | 454 | 2 F75580 | probable sugar tra |
| 4 | 358.5 | 25.1 | 422 | 2 G72234 | hypothetical prote |
| 5 | 348 | 24.4 | 400 | 2 C69757 | transporter homolo |
| 6 | 342 | 23.9 | 742 | 1 A43344 | synaptic vesicle p |
| 7 | 341 | 23.9 | 724 | 2 I50531 | transmembrane tran |
| 8 | 338 | 23.7 | 742 | 2 S27263 | synaptic vesicle p |
| 9 | 320 | 22.4 | 455 | 2 B83213 | probable MFS trans |
| 10 | 320 | 22.4 | 683 | 2 S34961 | synaptic vesicle p |
| 11 | 319 | 22.3 | 452 | 2 A85852 | probable transport |
| 12 | 281 | 19.7 | 593 | 2 JC4884 | organic cation tra |
| 13 | 278 | 19.5 | 536 | 2 S50862 | organic cation tra |
| 14 | 267.5 | 18.7 | 448 | 2 G83616 | 4-hydroxybenzoate |
| 15 | 265 | 18.5 | 446 | 2 B83033 | probable MFS trans |
| 16 | 264 | 18.5 | 448 | 2 H83335 | probable MFS trans |
| 17 | 244.5 | 17.1 | 444 | 2 E83033 | probable MFS trans |
| 18 | 242.5 | 17.0 | 451 | 2 S66008 | transport protein |
| 19 | 237.5 | 16.6 | 401 | 2 E69501 | sugar transporter |
| 20 | 231 | 16.2 | 557 | 2 JW0089 | organic cation tra |
| 21 | 222 | 15.5 | 423 | 2 S74046 | probable sugar tra |
| 22 | 215 | 15.0 | 557 | 2 JE0346 | high-affinity carn |
| 23 | 213 | 14.9 | 576 | 2 T22509 | hypothetical prote |
| 24 | 210 | 14.7 | 450 | 2 C83394 | probable MFS trans |
| 25 | 208 | 14.6 | 457 | 2 E70070 | metabolite transpo |
| 26 | 205.5 | 14.4 | 464 | 2 F69587 | L-arabinose transp |
| 27 | 202.5 | 14.4 | 469 | 2 G65058 | hypothetical prote |
| 28 | 202.5 | 14.2 | 401 | 2 E85927 | partial probable t |
| 29 | 202.5 | 14.2 | 473 | 2 G69789 | sugar transporter |

| | | | | | |
|----|-------|------|-----|----------|--------------------|
| 30 | 200 | 14.0 | 522 | 2 S12042 | glucose transport |
| 31 | 200 | 14.0 | 522 | 2 E86246 | glucose transporte |
| 32 | 199 | 13.9 | 419 | 2 E69888 | metabolite transpo |
| 33 | 198.5 | 13.9 | 423 | 2 A83383 | probable MFS trans |
| 34 | 198 | 13.9 | 422 | 2 C70518 | probable nanT prot |
| 35 | 197 | 13.8 | 523 | 2 S25015 | monosaccharide tra |
| 36 | 194 | 13.6 | 409 | 2 T47026 | hypothetical prote |
| 37 | 193 | 13.5 | 443 | 2 E64725 | yaaU protein - Esc |
| 38 | 193 | 13.5 | 557 | 2 T38125 | myo-inositol trans |
| 39 | 192 | 13.4 | 443 | 2 H85485 | probable transport |
| 40 | 191.5 | 13.4 | 407 | 2 C64167 | hypothetical prote |
| 41 | 191 | 13.4 | 508 | 2 T05156 | probable glucose t |
| 42 | 186 | 13.0 | 418 | 2 A64763 | probable transport |
| 43 | 186 | 13.0 | 418 | 2 D85530 | probable transport |
| 44 | 186 | 13.0 | 516 | 2 T12199 | monosaccharid tran |
| 45 | 184.5 | 12.9 | 575 | 2 T43400 | myo-inositol trans |

ALIGNMENTS

RESULT 1
T23190
hypothetical protein ZK637.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
C:Accession: T23190
R:Craxton, M.
submitted to the EMBL, Data Library, April 1993
A:Reference number: Z19704
A:Accession: T23190
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-529 <WIL>
A:Cross-references: EMBL:Z22175; PDB:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1
C:Genetics:
A:Gene: CESP:ZK637.1
A:Map position: 3
A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 58.2%; Score 831.5; DB 2; Length 529;
Best Local Similarity 58.0%; Pred. No. 3.8e-64;
Matches 160; Conservative 43; Mismatches 70; Indels 3; Gaps 1;
QY 1 EDVAEAIQFGKQWKLKSLVTGLAWADAMEMMLISLILAPQLHCENRIPSMOVALTTSVVF 60
Db 70 DEAEALGFRFQULKSLITGMWADAMEMMLLSLISPALACEWGISVQQAALVTCVF 129
QY 61 VGMSSSTLWGNISDQYGRKTKLISVLMTLYYGLSAPFVYSWILVRLGLVGFQIGV 120
Db 130 SGMLSTTFWKCICDFGRKGLTSTLVACIMVIGSMSPHFYVLLFFRGLTGTGIGV 189
QY 121 PQSTLYAEFLPMKARAKICILLIEVFVFAIGTVFVFLAVFVMPISLGRWLLLSAPVLLL 180
Db 190 PQSTLYAEFLPTAQARACVVLLIESFAIGVFAEALFVFMESFGWRALMFLSSLPGLI 249
QY 181 FAVLCFWLPESARYDVLVSGNOEKATATKRIATENGAPMLGLKLTISRO---EDRCKMRD 237
Db 250 FAVASFWLPESARFDMASGHPERALETQAAMRVRVQLTGRVLSYSTKAGSESRGDIAN 309
QY 238 LFTPHFRWTTLLLMFINSNAPSYGLVLLTTELFQ 273
Db 310 LLSPDLRRTTLLMCIAWITAFSYGYGMVLTFTVLFQ 345

RESULT 2
S15786
glucose transport protein homolog - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C:Date: 06-Jan-1995 #sequence_revision 05-May-2000 #text_change 05-May-2000

C:Accession: S15786
R:Craxton, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; Hall, Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.
submitted to the EMBL Data Library, May 1991
A:Reference number: S15786

A:Accession: S15786
A:Molecule type: DNA
A:Residues: 1-300 <CRA>
A:Cross-references: EMBL:Z11115
C:Genetics:
A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 49.5%; Score 707; DB 2; Length 300;
Best Local Similarity 59.0%; Pred. No. 1.1e-53;
Matches 134; Conservative 37; Mismatches 56; Indels 0; Gaps 0;

QY 1 EDVAEATGCKFKWKLVSITGLAWMADAMEMMILSILAPOLHCEWRLPSPVOALLTSVVF 60
DB 70 DEAVEALGFRGFKUSILTGMAWADAMEMMILSILPALACEWGISVQOQALVITCVF 129
QY 61 VGMSSSTLWGNISDOYGRKTKLISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGGV 120
DB 130 SGMMLSSTFWGKICDRGRRKGLTFTLVACIMGVISGMSPHFVYVLLFRGLTGFIGGV 189
QY 121 POSVTLYAEPLPKAKACILLIEVFWAIGTVFEVVLAVFVMPSLGWRWLLILSAVPLL 180
DB 190 POSVTLYAEPLTAQAKCVLIESFWAIGVFEALLAVFVMSFGWRALMFLSSILPLGI 249
QY 181 FAVLCFWLPESARYDVLSGNQEKAIAITLKRATENGAPMPLGKLIIS 227
DB 250 FAVASFWLPESARFDMASGHPERALETLQNAARMNRVQLPTGRVSS 296

RESULT 3
F75580
probable sugar transporter - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75580
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75580
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-454 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12486.1; PID:g646078
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0271
A:Map position: 2
C:Superfamily: yaaU protein

Query Match 31.2%; Score 446.5; DB 2; Length 454;
Best Local Similarity 35.1%; Pred. No. 4.9e-31;
Matches 94; Conservative 52; Mismatches 115; Indels 7; Gaps 3;

QY 4 VEAIGFGKFWKLVSITGLAWMADAMEMMILSILAPOLHCEWRLPSPVOA--LLTSVVFV 61
DB 19 LDDLGLGRFQWKLKLAICGLTWAADAMEVLMLGFALPGISAFAELPKGSPAAATMLLTATFA 78
QY 62 GMMSSSTLWGNISDOYGRKTKLISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG-V 120
DB 79 GMLFGAWFWGLADRVGRRSVFUTLVALGVVFGLAGALAPTLFWLVARFLTGFAGGTL 138
QY 121 POSVTLYAEPLPKAKACILLIEVFWAIGTVFEVVLAVFV----MPSLGRWLLILSAV 176

DB 139 PVDYSMMAEFVPTAWRGRFLVYLESEFVAVCTVVVAALAWWVSTAFAPAEGRWLLGLAAL 198
QY 177 PLLLFAVLFCWLPESARYDVLSONQEKAIAITLKRATENGAPMPLGKLIISRQEDRKM 236
DB 199 PGLVGLIARIGIPDSRSLARGEQAARAAQKVAQANGCTLPAAPLAHPPEOPPRVSPA 258
QY 237 DLFTPHFRWTTLLLWFIWFSNAFYYGL 264
DB 259 QLFGRVGLARTPLLMVTWFLSLGYGJ 286

RESULT 4
G72234
hypothetical protein TM1603 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: G72234
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Hatt, D.H.; H Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome A:Reference number: A72200; MUID:99287316
A:Accession: G72234
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-422 <ARN>
A:Cross-references: GB:AE001804; GB:AE000512; NID:g4982160; PIDN:AAD36670.1; PID:g4 A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1603
C:Superfamily: hypothetical protein H11104

Query Match 25.1%; Score 358.5; DB 2; Length 422;
Best Local Similarity 32.0%; Pred. No. 1.7e-23;
Matches 89; Conservative 55; Mismatches 111; Indels 23; Gaps 7;

QY 1 EDVAEATGCKFKWKLVSITGLAWMADAMEMMILSILAPOLHCEWRLPSPVOALLTSVVF 60
DB 4 DEIVEKYVDKRTORRFLILTSIAMDFDAGVLLSFVLPVIVKEWNLTSQTGATIASATF 63
QY 61 VGMSSSTLWGNISDOYGRKTKLISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGGV 120
DB 64 LGMLFGALSFGVADLLGRKVSNLFFIVTITFTLSGSSSFETLLVRLGSLGFGYGL 123
QY 121 POSVTLY-AEFLPKAKACILLIEVFWAIGTVFEVVLAVFVMPSLGWRWLLILSAVPLL 179
DB 124 MPSFNAYLAETIRLGRYLVILLESSWAVGSILIGLFAVNVLPN--WRWVWFISIG-Y 180
QY 180 LFAVLFCWLPESARYDVLSONQEKAIAITL-KRIATENGAP----MPLGKLIISRQEDRGK 234
DB 181 LFPVFLRMPETPKYAFKLGKKEALERSLGRVEEVELPKKVKPILALL-----KREH 235
QY 235 MRDLTPHFRWTTLLLWFIWFSNAFYYGLVLLTTELF 272
DB 236 LKD-----TVVIWIAWFWVSVYYALFTWAPRIF 264

RESULT 5
C69757
transporter homolog yceI - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: C69757
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; B C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; G iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardir A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; M

A:Accession number: F008440
 A:Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAH12089.1; PID:g2632581
 A:Experimental source: strain 168
 C:Genetics:
 C:Gene: yceI
 C:Superfamily: hypothetical protein H1104

| | | | | |
|-----------------------|------------------|--------------------|--------|-------------|
| Query Match | 23.9% | Score 342; | DB 1; | Length 742; |
| Best Local Similarity | 27.4% | Pred. No. 8.3e-22; | | |
| Matches 86: | Conservative 55; | Mismatches 125; | Indels | |

| | | | |
|----|-----|---|-----|
| Qy | 1 | EDAVEAIGFGKFQWKL SVLTCLAWADAMEMMILS ILAPQLHC EWRLPSMQVALLTSWF | 60 |
| | | : : : : : : : : : | |
| Db | 153 | ETILRECGHGRFQNTLYFVLGLALMADGVEVFVGFLP SAEKDMC LSDSNKGMLGLIVY | 212 |

Db 153 ETILRECGHGRFQWTLTYFVLGLALMADGVEFVVGVLPSPAEKDMCLSDSNKGMGLIVY 212

Qv 61 VGMSSSTI WGNISDOYGRKGTGJKTSVWTLVYGTLSAFAPVYSWTLVJRGIVGEGJGG- 119

[illegible]

Db 213 LGMVGAFLWGLADRLGRQRQLLSLSVNSVFAFFSFVQYGTFLFCRLLSGVIGGS 272

QY 120 VPQSVTLYAEELPMKARAKCILLIEVFWAIGTVFEVVLAVFVMPSLG----- 166

Db 273 IPIVFSYSEFLAQEKGREHLSWLCMFWMIGGVYAAAMAWAII PHYGWSFQMCSAYQFHS 332

Qv 167 WPWIIITISAVBII I FAVI CEMI DESAPYDVI SCNOEKATATI KPIATEN-----CAP---M 219

Q7 |
|07|WWWEIISARVIFUUEUTAVCTWFEESANIDVLSSUNGERATATUKKATIEN|URF
||:::||:
||:::||:

Db 333 WRVFLVCAFPVSFAIGALTTPESPFFFLENGKHD EAMVLKQVHDTNMRAGHPRVF 392

QY 220 PLGKLIISROEDR-----GKMRDLFTPHFRWTTLLWF 252

Db 393 SVTHIKTHQEDLIEIQSDTGTWYQRWGVRALSGLGGQVWGNFLSCFSPYRRITLMMMG 452

On 252 INFORMATION

QY 233 WFSNAPSIGLVL 200
:||: :||||| :

Db 453 VWFTMSFSYYGLTV 466

RESULT 7
I50531

transmembrane transporter - electric ray (*Discopyge ommata*)

C:Species: Discospyge omnata
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999

C;Accession: I50531
D;Biodra O C : Knowles D : Bucklow W M

R; Blinova, P.S.; Nowies, K.; Buckley, K.M.
Gene 137, 299-302, 1994

A;Title: Conservation of the amino acid sequence of SV2, a transmembrane t
A;Reference number: I50531; MIMD:94131301

A: Accession: 150531
 A: Reference Number: 150531, N01D: 34131501

A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

A;Residues: 1-724 <BIN>
A_Cross-ref:06988888. CD:J23403. NID:472816. PIDN:MAM0025 1. BID:472817

A: Cross-references: GB:L23403; NID:g4/2816; PIDN:AAA49235.1; PID:g4/2817
C: Superfamily: synaptic vesicle protein SV2

Query Match 23.9%; Score 341; DB 2; Length 724;

Best Local Similarity 28.3%; pred. No. 9.8e-22;
Matches 89: Conservative 54: Mismatches 112: Indels 60: Gaps

[illegible]

QY 4 VEALGKFKWLSVLTGLAWMADAMEMMILSILAPQLHCEWRLPSPWQVALLTSWFVGM 63
:: | :|:| | :: | :|:| | :|:| | :|:| | :|:| |

Db 139 IQEGHGRFQWALFLVLGLSLMADGVEFVVGFLPSAETDMCVENSNSGWLGSIVYLG 198

Qy 64 MSSSTLWGNISDQYGRKTGLKISVLWTLYYGILSAFAPVYSWILVRLGLVGFIGG-VPQ 122

199 MICAFFENGSLADKMKPRPTLIICMSTNCFEFAFISSFVOCVYSIFIECFEEFACEGICCAVDV 258

22 100 MEDICAL WOODLANDS IN THE EAST VALLEY OF THE GREAT OREGON RIVER

```

QY      123 SVTLYAEFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVMPSLG-----WRW 169

```

— — — — —

Qy 61 VOMSGSSTLWGNISDQYGRKTLGISVLWTLYYGILSFAPVYSWILVRLGVGFIGG- 119
 :
Db 156 LCMAGAFITGLGLADLGRKKVLSMSLAINFASFSLSSFVGQYGAFLFCRULISGIGGS 215
 :
Qy 120 VPQSVTLYAEFLPKAKACILLIEVFVAIGTVFEVVLAVFMPISLG----- 166
 :
Db 216 LPIVEAYSEFLSRKEGRHEHLSWGIFWMGTGGIYASAMAWSIIPHYGMGSMGTNYHFHS 275
 :
Qy 167 WRLLILLSAVPLLLFAVLFCWLPEARSVDVLSNGOEKAIALTKRIATEN---CAPMPLG 222
 :
Db 276 WRVFVICALPATYSMVALKFPESPRLFLEMGKHDEAMMLIKQVHDHTNRAKGTP--E 332
 :
Qy 223 KLIISRQEDRGKMRDLF-----TPIFRW-----TTLL 249
 :
Db 333 KVFTVSHLIKTPKWDEFIEIOSSGTGWYQRWLVRFTFKQVWDNALCYVMGPYRMNTLI 392
 :
Qy 250 LWTFWFSNAFSYYGLVL 266
 :
Db 393 LAVWNFTMALSYGYGLTV 409

RESULT 11
A85852
probable transporter Z3394 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A85852
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamoussis, K.; Apodaca
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206351
A:Accession: A85852
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <STO>
A:Cross-references: GB:AE005174; MID:g12516456; PIDN:AAGS7277.1; GSPDB:GN00145; UMWG:Z3394
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3394

| | | | | |
|-----------------------|-------|--------------------|-------|-----------------|
| Query Match | 22.3% | Score 319; | DB 2; | Length 452; |
| Best Local Similarity | 29.0% | Pred. No. 4.8e-20; | | |
| Matches | 76; | Conservative | 58; | Mismatches 124; |
| | | | | Indels 4; |
| | | | | Gaps 4; |

QY 4 VEATGFGKFWKLSVLTGLAWMADAMMMILSLAPQLHCEWRLPWSQVALLTSVVEVGM 63
: : | | | | : : : : | | : : | : : : : :
Db 11 IDAAPVKMQWRVITCCFLVYMLDGDFTAAIGFIAPDINTHWQITAGDLSPLFGAGLLXL 70

Qy 64 MSSSTLWGNISDQYGRKTKGLKISVLWLTYYGILSAFAPVYSWILVRLGLVGFGIGG-VPQ 122

Db 71 TAGALLCGPLSDRFGKRVRIELCVFLTFGALSLASAFSPDQLTVFLRFLTGLGLGGAMPN 130

| | | |
|----|-----|---|
| QY | 123 | SVTLYAEEFLPMKAKACILLIEVFMAIGTVFEVVLAVFVMPSLGWRWLLLSAV - PLLLF 181 ::: : : : : : : : : : : : : |
| Db | 131 | TITMTSEYLPAARRRGALVTLMFCFGTILGSAGGVISQAQLPVPVGHWGHGLLVGGVLPMLF 190 ::: : : : : : : : : : : : : |

QY 182 AVLCEWLPESARYDVLSGOEKAIATLKRATENGAPMPLGKLIISRQEDRGKMRDLFT 240
 191 VALLVLPESRWQYRRLQPOAVIAKTVSATREYVDTHFYLESASVTGKSIROLFM 249
 Db

| | | | |
|----|-----|---|-----|
| Qy | 241 | P H F R W T T L L L W F I W F S N A F S Y Y | 262 |
| | | : : : : | |
| Db | 250 | G R O L P I T L M L W V F F M S L L I Y | 271 |

RESULT 12
JC4884

organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C:Accession: J04884

R;Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.
Biochem. Biophys. Res. Commun. 224, 500-507, 1996
A;Title: cDNA cloning and functional expression of a novel rat kidney organic cation
A;Reference number: JC4884; MUID:96295517
A;Accession: JC4884
A;Molecule type: mRNA
A;Residues: 1-593 <OKU>
A;Cross-references: DDBJ:D83044; NID:gl502282; PIDN:BA11754.1; PID:d1012421; PID:gl15
A;Experimental source: kidney
C;Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 19.7%; Score 281; DB 2; Length 593;
Best Local Similarity 33.0%; Pred. No. 1.2e-16;

[illegible][illegible]

Qy 161 VMPSLGWRWLLILSAVPLLLFAVLCFW-LPESARYDVLSCNQEKAIATIKRIATENAGPM 219
|::: |||| : | : | : || : : ||: : || : || :

220 PLGKLIISQEDRCK-----MRDLF-TPHFRWTTLLLWFIFWNAFSYGLVL 266

00 313 PVSQNLIFDEDAQNNKAPSLDLVRKIPQIRARHILILMINWFTSSVLIQGLIM 36 /

RESULT 13

350662
organic cation transport protein OCT1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Nov-1999

C/Accession: S50862; S78533; I58089
R/Gruendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.
Nature 372, 549-552, 1994
A>Title: Drug excretion mediated by a new prototype of polyspecific transp

A; Accession: S50852
A; Reference number: 158089; MUID: 95082907
A; Molecule type: mRNA
A; Residues: 1-556 <GRU>

A;Experimental source: kidney
A;Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622
R;Gorboulev, V.G.
submitted to the EMBL Data Library, January 1995

A; Reference number: S78533
A; Accession: S78533
A; Molecule type: mRNA
A; Residues: 1-342, 'N', 344-556 <GOR>

A;C:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g6336622
 C;Keywords: glycoprotein; phosphoprotein; transmembrane protein
 F;20-46/Domain: transmembrane #status predicted <TM1>
 F;154-171/Domain: transmembrane #status predicted <TM2>

| Accession | Domain | transmembrane | #status | predicted |
|-----------|--------|---------------|---------|-----------------|
| F_178-197 | Domain | transmembrane | #status | predicted <TM3> |
| F_243-260 | Domain | transmembrane | #status | predicted <TM4> |
| F_267-283 | Domain | transmembrane | #status | predicted <TM5> |
| F_350-366 | Domain | transmembrane | #status | predicted <TM6> |

| | | | | |
|-------------------|---------------|---------|-----------|--------|
| F;380-398/Domain: | transmembrane | #status | predicted | <TM7> |
| F;406-425/Domain: | transmembrane | #status | predicted | <TM8> |
| F;435-452/Domain: | transmembrane | #status | predicted | <TM9> |
| F;469-485/Domain: | transmembrane | #status | predicted | <TM10> |

F:494-514/Domain: transmembrane #status predicted <TM1>
F:71-97,113,432/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:286,292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #s
F:296,343,550/Binding site: phosphate(Thr) (covalent) (by protein kinase C)
F:296,343,550/Binding site: phosphate(Thr) (covalent) (by protein kinase C)

Query Match 19.5%; Score 278; DB 2; Length 556;

Best Local Similarity 32.0%; Pred. No. 2e-16;
Matches 72; Conservative 48; Mismatches 95; Indels 10; Gaps 5;

QY 49 SMOVALLTSVVFVGMSSSTLWGNISDOYGRKTKISVLWTLTYGILSAFAPVYSWILV 108
Db 146 AKAVDLFQSCVNLGFFLGVGYIADRGRKLCLLVTLTVSVGVLTAVAPDYTSMLL 205
QY 109 LRGLVGF-GIGGVQPSVTLYAEFLPMKARAKCILLIEVFAIGTVFEVVLAVFVMPISGW 167
Db 206 FRLLQGMVSKGWSVGYTLITEFVSGYRRTAILYQMAFTVGLVGLAGVA-YAIPD--W 262
QY 168 RWLLILSAVPLLLFAVLGFWLPESARYDVLGNGQEKAIATLKRIATENGAPMPLGLKLIIS 227
Db 263 RWLQLAVSLPTFLFLYYWFVSPESRWLLSQKRTTRAVRIMEQIAQKNGKVPADLKMIC 322
QY 228 RQDRGKMR-----DLF-TPHFRWTTLLLWFIWFSNAFSYGLVL 266
Db 323 LEEDASEKRSFADLFRPTLRKHTVILMYLWFSCAVLVYQGLIM 367
RESULT 14
G83616
4-hydroxybenzoate transporter PcaK PA0235 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83616
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:20437337
A:Accession: G83616
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <STO>
A:Cross-references: GB:AE004461; GB:AE004091; NID:99946066; PIDN:AAG03624.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: pcaK; PA0235

Query Match 18.7%; Score 267.5; DB 2; Length 448;
Best Local Similarity 29.2%; Pred. No. 1.3e-15;
Matches 81; Conservative 54; Mismatches 125; Indels 17; Gaps 8;

QY 4 VEAIGFGKQWKLSVLTGLAMMADAMEMMILSILAPOLHCEWRLPSWQVALLTSVVPVGM 63
Db 18 INAQPLSPYQWRIVLLCFILVFLDGLDTAAMGFIAPALTQIWGIDRASLGPVMSAALIGM 77
QY 64 MSSSTLWGNISDOYGRKTKISVLWTLTYGILSAFAPVYSWILVRLGLVGFPGIG-GVPPQ 122
Db 78 VFGALGSGPLADRYGRKLVVAAVFLGFLSLASAYSTNVEQLALRFLTLGLGLGAAMPN 137
QY 123 SVTLXAEFLPMKARAKCILLIEVF--WAIGTVFEVVLAVFVMPISGWRWLLTSLA-VPLL 179
Db 138 ATLLSEYTP--ERLKSLLVTSFMFCGFLNMACCGFVSAKLIPLFGWHSLLLDGGLPLV 195
QY 180 LFVAVLCFWLPESARYDVL-SGNQEKAIATLKRIATENGAPMPLGLKLIISROEDRGKMRDL 238
Db 196 LAVVLLFRLPESARYLVVVRNRSERVROVLAPIAP--AQVALARSFHVPEQQTVAARNV 252
QY 239 FTPHFRWT-----TLLLWFIWFSNAFSYGLVLLTTEL 271
Db 253 FAVIFSGTYSAGTLLLLMLTYFMGLVIVY---LLTSWL 286

RESULT 15
B83033
Probable MFS transporter PA4900 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83033

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A:Reference number: A82950; MUID:20437337
A:Accession: B83033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <STO>
A:Cross-references: GB:AE004903; GB:AE004091; NID:99951173; PIDN:AAG08285.1; GSPDB:B
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4900
C:Superfamily: yaaU protein

Query Match 18.5%; Score 265; DB 2; Length 446;
Best Local Similarity 25.8%; Pred. No. 2.1e-15;
Matches 70; Conservative 58; Mismatches 123; Indels 20; Gaps 6;

QY 4 VEAIGFGKQWKLSVLTGLAW-----MADAMEMMILSILAPOLHCEWRLPSWQVALLTSV 58
Db 10 IDAARFNRTHWLI-----LGMGCFIMLFDGYDMVITYGSVVPRLMQEQLSPVQAGTLGSC 64
QY 59 VFVGMSSSTLWGNISDOYGRKTKISVLWTLTYGILSAFAPVYS---WILVLRGLVG 114
Db 65 ALFGMLFGGTTLLAPLADRFGRRR---LVIAITTLASLAFLTGHDPLGACGRFFTG 120
QY 115 FGIGG-VPOSVTLYAEFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVMPISGWRWLLIL 173
Db 121 LALGALVPSAINLISEFAPAGRRTSLVTVMISAFYSYSGAVLSALLAIAIAMPWGQSVFYV 180
QY 174 SAVPLLLFAVLGFWLPESARYDVLGNGQEKAIATLKRIATENGAPMPLGLKLIISROEDRG 233
Db 181 AVLPLVAVPLMLRWLPESARFLELKGRRRAELDALLRKVDYDYP-RPGAERANAVAAEAPSG 239
QY 234 KMRDLFTPHFRWTTLLLMFIMFSNAFSYGL 264
Db 240 RVAQLFEGRQAVGTLLLVAVAFAMCMLMSYGL 270

Search completed: March 13, 2002, 12:41:39
Job time: 85 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:42:52 ; Search time 17.05 seconds
(without alignments)
587.067 Million cell updates/sec

Title: US-09-911-667A-4
Perfect score: 1429
Sequence: 1 EDVAIGAIGKQWKLKSLVT.....WFSNAPSYGLVLLTTELTV 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 831.5 | 58.2 | 529 | 1 Y0UL_CAEEL | P30638 caenorhabdi |
| 2 | 348 | 24.4 | 400 | 1 YCEI_BACSU | O34691 bacillus su |
| 3 | 341 | 23.9 | 742 | 1 SYV2_RAT | O02563 rattus norv |
| 4 | 282.5 | 19.8 | 448 | 1 PCAK_PSEPU | O31955 pseudomonas |
| 5 | 260.5 | 18.2 | 457 | 1 PCAK_ACICA | O43975 acinetobact |
| 6 | 242.5 | 17.0 | 451 | 1 YTAJ_BACSU | P37514 bacillus su |
| 7 | 238 | 16.7 | 466 | 1 BENK_ACICA | O30513 acinetobact |
| 8 | 231 | 16.2 | 557 | 1 OCN2_HUMAN | O76082 homo sapien |
| 9 | 227.5 | 15.9 | 413 | 1 MUCK_ACICA | P94131 acinetobact |
| 10 | 219 | 15.3 | 473 | 1 PHDK_NOSK | O24723 nocardioid |
| 11 | 219 | 15.3 | 557 | 1 OCN2_MOUSE | O92088 mus musculu |
| 12 | 219 | 15.3 | 557 | 1 OCN2_RAT | O70594 rattus norv |
| 13 | 205.5 | 14.4 | 445 | 1 YGCS_ECOLI | O46909 escherichia |
| 14 | 203.5 | 14.2 | 428 | 1 MMLH_ALCEU | O51798 alcaligenes |
| 15 | 200 | 14.0 | 522 | 1 STP1_ARATH | P23586 arabidopsis |
| 16 | 193 | 13.5 | 443 | 1 YAAU_ECOLI | P31679 escherichia |
| 17 | 193 | 13.5 | 557 | 1 ITR2_SCHPO | P87110 schizosacch |
| 18 | 191.5 | 13.4 | 407 | 1 YB04_HAEIN | K71365 haemophilus |
| 19 | 186 | 13.0 | 403 | 1 MHPT_ECOLI | P71589 escherichia |
| 20 | 184.5 | 12.9 | 575 | 1 ITRI_SCHPO | Q10286 schizosacch |
| 21 | 184 | 12.9 | 547 | 1 GTRI_LEIDO | Q01440 leishmania |
| 22 | 181 | 12.7 | 523 | 1 STC_RICCO | Q41144 ricinus com |
| 23 | 178 | 12.5 | 457 | 1 XYL1_LACBR | O52733 lactobacill |
| 24 | 177 | 12.4 | 496 | 1 NANT_ECOLI | P41036 escherichia |
| 25 | 175.5 | 12.3 | 751 | 1 YLX5_CAEEL | P46501 caenorhabdi |
| 26 | 171.5 | 12.0 | 461 | 1 YFBC_BACSU | P46333 bacillus su |
| 27 | 167.5 | 11.7 | 482 | 1 CSIC_BACSU | P54723 bacillus su |
| 28 | 166.5 | 11.7 | 468 | 1 GLCP_SYNY3 | P15729 synecocyst |
| 29 | 162 | 11.3 | 540 | 1 HUP2_CHLKE | Q39524 chlorella k |
| 30 | 161 | 11.3 | 491 | 1 XYLE_ECOLI | P09098 escherichia |
| 31 | 158.5 | 11.1 | 522 | 1 GTR2_RAT | P12336 rattus norv |
| 32 | 158 | 11.1 | 534 | 1 HUP3_CHLKE | Q39525 chlorella k |
| 33 | 157 | 11.0 | 435 | 1 YTI3_CAEEL | Q10917 caenorhabdi |

RESULT 1

| ID | Y0UL_CAEEL | STANDARD; | PRT; | 529 AA. |
|-----|--|-----------|------|---------|
| AC | P30638; O21101; | | | |
| DT | 01-APR-1993 (Rel. 25, Created) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | |
| DE | HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III. | | | |
| GN | ZK637.1. | | | |
| OS | Caenorhabditis elegans. | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; | | | |
| OC | Rhabditidae; Peloderinae; Caenorhabditis. | | | |
| OX | NCBI_TaxID=6239; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-BRISTOL N2; | | | |
| RX | MEDLINE=92168156; PubMed=1538779; | | | |
| RA | Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R., | | | |
| RA | Halloran N., Green P., Thierry-Mieg J., Oiu L., Dear S., Coulson A., | | | |
| RA | Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T., | | | |
| RA | Alnscough K., Waterston K.; | | | |
| RT | *The C. elegans genome sequencing project: a beginning.*; | | | |
| RL | Nature 356:37-41(1992). | | | |
| [2] | | | | |
| RP | REVISIONS. | | | |
| RC | STRAIN-BRISTOL N2; | | | |
| RA | Durbin R.; | | | |
| RL | Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE). | | | |
| CC | -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; Z11115; CAA77460.1; .. | | | |
| DR | EMBL; Z22175; CAA77460.1; JOINED. | | | |
| DR | EMBL; Z22175; CAA80131.1; .. | | | |
| DR | EMBL; Z11115; CAA80131.1; JOINED. | | | |
| DR | PIR; S15786; S15786. | | | |
| DR | WormPep; ZK637.1; CE066638 | | | |
| DR | InterPro; IPR003662; sub_trnsportr. | | | |
| DR | Pfam; PF00083; sugar_tr; 1. | | | |
| DR | PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG. | | | |
| DR | PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG. | | | |
| KW | Hypothetical protein; Transmembrane; Transport. | | | |
| FT | TRANSMEM 86 106 POTENTIAL. | | | |
| FT | TRANSMEM 122 142 POTENTIAL. | | | |
| FT | TRANSMEM 158 178 POTENTIAL. | | | |
| FT | TRANSMEM 230 258 POTENTIAL. | | | |
| FT | TRANSMEM 328 340 POTENTIAL. | | | |
| FT | TRANSMEM 373 393 POTENTIAL. | | | |

ALIGNMENTS

| | | | | | |
|----|-------|------|------|--------------|--------------------|
| 34 | 156.5 | 11.0 | 464 | 1 GALP_ECOLI | P37021 escherichia |
| 35 | 155 | 10.8 | 522 | 1 STA_RICCO | Q10710 ricinus com |
| 36 | 154.5 | 10.8 | 584 | 1 ITRI_YEAST | P30605 saccharomyc |
| 37 | 152 | 10.6 | 559 | 1 YAEC_SCHPO | Q09852 schizosacch |
| 38 | 150.5 | 10.5 | 523 | 1 GTR2_MOUSE | P14246 mus musculu |
| 39 | 149 | 10.4 | 612 | 1 ITR2_YEAST | P30606 saccharomyc |
| 40 | 148.5 | 10.4 | 1222 | 1 YMP3_CAEEL | Q10947 caenorhabdi |
| 41 | 148 | 10.4 | 472 | 1 EXUT_ECOLI | P42609 escherichia |
| 42 | 146.5 | 10.3 | 546 | 1 HXT0_YEAST | P43581 saccharomyc |
| 43 | 146.5 | 10.3 | 570 | 1 HXT1_YEAST | P32465 saccharomyc |
| 44 | 145.5 | 10.2 | 422 | 1 EXUT_BACSU | O34456 bacillus su |
| 45 | 145.5 | 10.2 | 524 | 1 GTR2_HUMAN | P11168 homo sapien |

```
FT  TRANSMEM  411  431  POTENTIAL.
FT  TRANSMEM  482  502  POTENTIAL.
SQ  SEQUENCE  529 AA;  58317 MW;  8D2FF4CBA15ECD2D CRC64;

Query Match
Best Local Similarity  58.2%;  Score 831.5;  DB 1;  Length 529;
Matches 160;  Conservative 43;  Mismatches 70;  Indels 3;  Gaps 1;

QY  1  EDVAEAGCKFKOWKLSVLTGLAWMADAMEMMILSLAPOLHCEWRLPSQVALLTSVVF 60
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  70  DEAVEALGFRGOLKLSLTGMAWADAMEMMILSLAPALACEWGISVQQAALVTCVF 129

QY  61  VGMSSSTLWNGISDQYGRKTKLKVSLWLTLYGILSAFAPVYSWTLVRLGLVFGIGGV 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  130  SGMMLSTFWGKICDFRGRKGLTFTLVACIMGVISGMSPHFYVLLFRLGLTGFVIGGV 189

QY  121  PQSVTLVLAFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVPSLGRWRLILLSAYPLLL 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  190  PQSVTLVLAFLPTAQRAKCVLLIESFWAIGAVFEALLAVFVMSFGWRLMFLSSPLGI 249

QY  181  FAVLCWLPESARYDVLGNOEKAIAIKRIATENGAPMLGKLIISRQ---EDRGKMRD 237
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  250  FAVASFWLPESARFDMASGHPERALFTLQAAARMNRVQLPTGRVLSSTRAGSRGDIAN 309

QY  238  LFTPHI.RWTLTLLWTFWISNAPSYYGLVLLTTLFQ 273
    |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  310  LLSPLDKRTTILICWIAITAFSYGMVLTFTVLQ 345

RESULT  2
YCEL_BACSU  STANDARD;  PRT;  400 AA.
AC  Q34691;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YCEI.
GN  YCEI.
OS  Bacillus subtilis.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/Staphylococcus group; Bacillus.
OX  NCBI_TaxID=1423;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RA  Kumano M., Tamakoshi A., Yamane K.;
RT  "A 32 kb nucleotide sequence from the region of the lincomycin-
RT  resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT  identification of the site of the lin-2 mutation.";
RL  Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC  -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AB000617; BAA22256.1; -.
DR  EMBL; Z99105; CAB12089.1; -.
DR  Subtilist; BG12773; ycel.
DR  InterPro; IPR003662; sub_trnsport.
DR  Pfam; PF00083; sugar_tr; 1.
DR  PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR  PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW  Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT  TRANSMEM  15  35  POTENTIAL.
FT  TRANSMEM  50  70  POTENTIAL.
FT  TRANSMEM  78  98  POTENTIAL.
```

```
FT  TRANSMEM  99  119  POTENTIAL.
FT  TRANSMEM  143  163  POTENTIAL.
FT  TRANSMEM  166  186  POTENTIAL.
FT  TRANSMEM  218  238  POTENTIAL.
FT  TRANSMEM  254  274  POTENTIAL.
FT  TRANSMEM  281  301  POTENTIAL.
FT  TRANSMEM  305  325  POTENTIAL.
FT  TRANSMEM  344  364  POTENTIAL.
FT  TRANSMEM  371  391  POTENTIAL.
SQ  SEQUENCE  400 AA;  43708 MW;  E0AE0CEE5DD27395 CRC64;

Query Match
Best Local Similarity  24.4%;  Score 348;  DB 1;  Length 400;
Matches 87;  Conservative 51;  Mismatches 86;  Indels 34;  Gaps 6;

QY  13  QNKLSVLTGLAWMADAMEMMILSLAPOLHCEWRLPSQVALLTSVVFVGMSSSTLWGN 72
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  9  QRKLLGVAGLGLFDAMDVGILSFIILAHVWNLSPPEMKWIGSVNSIGMAAGAFLEGL 68

QY  73  ISDQYGRKTKLKVSLWLTLYGII---LSAFAPVYSWTLVRLGLVFGIGG-VPOSVTLYA 128
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  69  LADRIGRK---KVFITLLCFSGISGSAFVTSLSAFLILRFVIGMGLGELPVASTLVS 125

QY  129  EFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVMPSLGRWRLILLSAYPLLLFAVLCFWL 188
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  126  EAVVPEKRGRIVLLSFVAVGWLAAALISYFVPSFGWQAALLLTALTAFAYALYLRSL 185

QY  189  PESARYDVLGNOEKAIAIKRIATENGAPMLGKLIISKQEDRGKMRDLTFPHFRWTL 248
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  186  POSPRYSLSAK-----KRSMMWEN-----VKSWAREQYIR-----PTV 218

QY  249  LLWFIMFSNAPSYYGLVL 266
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  219  MLSIWFVCFVFSYGMFL 236

RESULT  3
SYV2_RAT  STANDARD;  PRT;  742 AA.
AC  Q02563;
DT  01-FEB-1995 (Rel. 31, Created)
DT  01-FEB-1995 (Rel. 31, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  SYNAPTIC VESICLE PROTEIN 2 (SV2).
GN  SV2.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
RC  TISSUE=Brain;
RX  MEDLINE=92390722; Pubmed=1519064;
RA  Rajjallah S.M., Peterson K., Shingal R., Scheller R.H.;
RT  "SV2, a brain synaptic vesicle protein homologous to bacterial
RT  transporters.";
RL  Science 257:1271-1273(1992).
CC  -!- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER
CC  TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.
CC  -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.
CC  -!- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND
CC  SPINAL CORD.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L05435; AAA42188.1; -.
```

DR Pfam: PF00083; sugar_tr; 1.
KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;
KW Transmembrane.
FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 164 182 POTENTIAL.
FT DOMAIN 183 204 LUMENAL (POTENTIAL).
FT TRANSSEM 205 225 POTENTIAL.
FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 243 260 POTENTIAL.
FT DOMAIN 261 262 LUMENAL (POTENTIAL).
FT TRANSSEM 263 281 POTENTIAL.
FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 295 322 POTENTIAL.
FT DOMAIN 323 334 LUMENAL (POTENTIAL).
FT TRANSSEM 335 355 POTENTIAL.
FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 446 469 POTENTIAL.
FT DOMAIN 470 592 LUMENAL (POTENTIAL).
FT TRANSSEM 593 611 POTENTIAL.
FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 627 647 POTENTIAL.
FT DOMAIN 648 649 LUMENAL (POTENTIAL).
FT TRANSSEM 650 669 POTENTIAL.
FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 695 711 POTENTIAL.
FT DOMAIN 712 712 LUMENAL (POTENTIAL).
FT TRANSSEM 713 731 POTENTIAL.
FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 742 AA; 82705 MW; 565DE7EF2929D5DB CRC64;

Query Match 23.9%; Score 341; DB 1; Length 742;
Best Local Similarity 27.4%; Pred. No. 2,7e-18;
Matches 86; Conservative 55; Mismatches 125; Indels 48; Gaps 5;

QY 1 EDVAEATGCKFKQKLVLTGLAWMADAMEMMILSIAPQLHCEWRLPSWOVALLTSVVF 60
Db 153 ETIRECGHGFRQWTLTVFLGLALMADGVEVVFVGLPSAEKDMCLSDSNKGMGLIVY 212
QY 61 VGMSSSTLMGNSIDGKRTGLKISVLWTLYGILSAFAPVSWILVLRGLVGFGLGG- 119
Db 213 LGMVGAFLAGLADLRGRQCLLSISVNSVFAFFSVQGYCTFLFCRLLSVGVGG 272
QY 120 VPQSVTLVAFELPKAKAKILLIEVFWAIGTVFVVLAVFVMPISG----- 166
Db 273 IPVFSYSEFLAQEKRGHLSWLCMFMMIGGVYAAAMAAIIPHYGWSFGMSAYOFHS 332
QY 167 WRLLILSAVPLLLFVLVCFWLPESARYDVLSGNOEKAIAIATLKRIATEN---GAP--M 219
Db 333 WRVFLVFAFSPVFAIGALTQSPSPRFFLENGKHDEAMVVLKQVHDTNRAKGHPERV 392
QY 220 PLGLLIISRQED-----GKMRDLFTPHFRWTTLLWF 252
Db 393 SVTHIKTIHOEDELIEIQTSDGTGTQVQRVRLSLGQGVWNGFLSCFSPRYRITLAMMG 452
QY 253 IWFNSAFSYGLVL 266
Db 453 VWFTMSFSYGLTV 466

RESULT 4
PCAK_PSEPU
ID PCAK_PSEPU STANDARD; PRT; 448 AA.
AC Q51955;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 4-HYDROXYBENZOATE TRANSPORTER.
GN PCAK.
OS Pseudomonas putida.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PRS2000;
RX MEDLINE=95050205; PubMed=7961399;
RA HARWOOD C.S., NICHOLS N.N., KIM M.-K., DITTY J.L., PARALES R.E.;
RT *Identification of the pcakK gene cluster from Pseudomonas putida:
RT involvement in chemotaxis, biodegradation, and transport of
RT 4-hydroxybenzoate*;
RL J. Bacteriol. 176:6479-6488(1994).
CC -!- FUNCTION: TRANSPORTER FOR 4-HYDROXYBENZOATE. ALSO REQUIRED FOR
CC CHEMOTAXIS TO AROMATIC ACIDS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U10895; AAA85137.1; -
CC InterPro: IPR003662; sub_transportr.
CC Pfam: PF00083; sugar_tr; 1.
CC PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
CC PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Transport; Transmembrane; Inner membrane.
FT TRANSSEM 31 51 POTENTIAL.
FT TRANSSEM 51 58 POTENTIAL.
FT TRANSSEM 58 88 POTENTIAL.
FT TRANSSEM 88 115 POTENTIAL.
FT TRANSSEM 115 140 POTENTIAL.
FT TRANSSEM 140 173 POTENTIAL.
FT TRANSSEM 173 205 POTENTIAL.
FT TRANSSEM 205 282 POTENTIAL.
FT TRANSSEM 282 322 POTENTIAL.
FT TRANSSEM 322 350 POTENTIAL.
FT TRANSSEM 350 372 POTENTIAL.
FT TRANSSEM 372 419 POTENTIAL.
FT TRANSSEM 419 442 POTENTIAL.
FT TRANSSEM 442 442 POTENTIAL.
SQ SEQUENCE 448 AA; 47176 MW; 6D51C143123E99BC CRC64;

Query Match 19.8%; Score 282.5; DB 1; Length 448;
Best Local Similarity 30.9%; Pred. No. 3.6e-14;
Matches 83; Conservative 54; Mismatches 117; Indels 15; Gaps 8;

QY 11 KFOWKLSVLTGLAWMADAMEMMILSIAPQLHCEWRLPSWOVALLTSVVFYGMSSSTLW 70
Db 25 RYQWRVLLCLFVLFDGLDGTAAAGFTAPALSQEWGIDRASLGPVMSAALIGMVFGALGS 84
QY 71 GNISDOYGRKTGLKISVLWTLYGILSAFAPVSWILVLRGLVGFGLGG- 129
Db 85 GPLADRFGRKGVLVGAVLVFGFGLASAYATNVDQLLVRLFTGLGLGAGHPNATLLSE 144
QY 130 FLPMKAKAKICILLIEVF--WAIGTVFVFLAVFVMPISLGMFWLLILSAV-PLLLFAVLGF 186
Db 145 YTP--ERLKSLLVTSMEFGNGLNAGGGFTISAKMIPAYGHSLLVIGVLPDLLALVLMV 202
QY 187 WLPESARYDVL-SGNOEKAIAIATLKRIATENGAPMPLKGLIISRQE---DRGKMRDLFTPH 242
Db 203 WLPESARFLVVRNRGTDKIRKTLSPAPIQVVA--EAGSFSVPEQKAVARSVFAVIESGT 260
QY 243 FRWTTLLIWFIFESNAFSGYGLVLLTTEL 271
Db 261 YGLGTMILLTLTYFMGLVIVY---LLTSWL 286

RESULT 5

| PCAK_ACICA | PCAK_ACICA | STANDARD; | PRT; | 457 AA. |
|------------|--|-----------|------|---------|
| ID | Q343975; | | | |
| AC | 15-JUL-1999 (Rel. 38, Created) | | | |
| AD | 15-JUL-1999 (Rel. 38, Last sequence update) | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | |
| DE | 4-HYDROXYBENZOATE TRANSPORTER. | | | |
| DN | PCAK. | | | |
| GN | Acinetobacter calcoaceticus. | | | |
| OS | Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; | | | |
| OC | Acinetobacter. | | | |
| OX | NCBI_TaxID=471; | | | |
| ON | [1] | | | |
| RN | SEQUENCE FROM N.A. | | | |
| RP | STRAIN=BD413 / ADP1; | | | |
| RC | MEDLINE=94341565; PubMed=8063101; | | | |
| RX | Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.-L., | | | |
| RA | Ornston L.N.; | | | |
| RT | "Contrasting patterns of evolutionary divergence within the | | | |
| RT | Acinetobacter calcoaceticus pca operon." | | | |
| RL | Gene 146:23-30(1994). | | | |
| CC | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE | | | |
| CC | (POTENTIAL). | | | |
| CC | -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; L05770; AAC37151.1; -- | | | |
| DR | InterPro; IPR003662; sub_transportr. | | | |
| DR | Pfam; PF00083; sugar_tr; 1. | | | |
| DR | PROSITE; PS00216; SUGAR_TRANSPORT_1; 1. | | | |
| DR | PROSITE; PS00217; SUGAR_TRANSPORT_2; 1. | | | |
| KW | Transport; Transmembrane; Inner membrane. | | | |
| FT | TRANSMEM 35 55 POTENTIAL. | | | |
| FT | TRANSMEM 73 93 POTENTIAL. | | | |
| FT | TRANSMEM 102 122 POTENTIAL. | | | |
| FT | TRANSMEM 129 149 POTENTIAL. | | | |
| FT | TRANSMEM 169 189 POTENTIAL. | | | |
| FT | TRANSMEM 192 212 POTENTIAL. | | | |
| FT | TRANSMEM 275 295 POTENTIAL. | | | |
| FT | TRANSMEM 311 331 POTENTIAL. | | | |
| FT | TRANSMEM 339 359 POTENTIAL. | | | |
| FT | TRANSMEM 365 385 POTENTIAL. | | | |
| FT | TRANSMEM 401 421 POTENTIAL. | | | |
| FT | TRANSMEM 427 447 POTENTIAL. | | | |
| SQ | SEQUENCE 457 AA; 49277 MW; 4F5B5F77361A1567 CRC64; | | | |

| Query Match | 18.2% | Score 260.5; | DB 1; | Length 457; |
|-----------------------|--------|---|-----------------|--------------------|
| Best Local Similarity | 25.8%; | Pred. No. 1.6e-12; | | |
| Matches | 70; | Conservative 64; | Mismatches 120; | Indels 17; Gaps 6; |
| QY | 11 | KFQNKLSVLGTGLAWMADAMEMMTLSILAPOLHCEWRLPSSQVALLTTSVFVGMSSSTFLW | 70 | |
| Db | 32 | RYQWLIATFLVFDVGDITDAAMGFIAPALADGWDRSQLGPMVAALGGMIIAGLV | 91 | |
| QY | 71 | GNISDQVGRGTGKISVLWLTLYGILSAPVYSWILVRLGLVGFIGV-GVPOSVTLVYAE | 129 | |
| Db | 92 | GPTADREGRKIVLSMSLNVFEGGTLCACAYSTNDSLIVIFRFTGIGLGAAMPNATTLFSE | 151 | |
| QY | 130 | FLPMKARAKICILIEVFATGTVFVVVAVFVPSLGWRLLTILSA-VPLLLFAVLFCWL | 188 | |
| Db | 152 | YCPARISSLVTCMFCGYNLGMAGIGGFISWLLIPAFGWHSLSFLILGGWAPILMLLVIFFL | 211 | |
| QY | 189 | PESARYDVLGS-NQEKAIATLKRIA-----TENGAPMLGKLIIISRQEDRCKMRDLFT | 240 | |
| Db | 212 | PESYRFLIVLGGKNTKKVRQYILSRAPKQGVGVTFTHVP-----BEKVEAGTKGVGFMLES | 267 | |

```

Qy      241  PHFRWTTLLWFIWFSNAFSYYGLVLLTTEL 271
      :  I:III :I      I      III: I
Db      268  AKYVKGTVLLWTVFMGLVMIV---LLTSWL 295

RESULT 6
YYAJ_BACSU STANDARD; PRT; 451 AA.
ID YYAJ_BACSU
AC P37514;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YYAJ.
GN YYAJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=96051385; PubMed=7584024;
RX Ogasawara N., Nakai S., Yoshikawa H.;
RA "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26185; BAA05214.1; -
CC DR EMBL; Z99124; CAB16121.1; -
CC DR Subtilist; BG10042; YYAJ.
CC DR InterPro; IPR003662; sub_trnsprtr.
CC PFam; PF00083; sugar_tr; 1.
CC DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE NEG.
CC DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE NEG.
CC KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
CC FT TRANSMEM 30 50 POTENTIAL.
CC FT TRANSMEM 61 81 POTENTIAL.
CC FT TRANSMEM 99 119 POTENTIAL.
CC FT TRANSMEM 128 148 POTENTIAL.
CC FT TRANSMEM 159 179 POTENTIAL.
CC FT TRANSMEM 183 203 POTENTIAL.
CC FT TRANSMEM 271 291 POTENTIAL.
CC FT TRANSMEM 306 326 POTENTIAL.
CC FT TRANSMEM 345 365 POTENTIAL.
CC FT TRANSMEM 397 417 POTENTIAL.
CC FT TRANSMEM 422 442 POTENTIAL.
CC SQ SEQUENCE 451 AA; 496672 MW; 4A00D3891C4D7D09 CRC64;

```

| | | | | | |
|----|-----------------------|--|--------------------|-------|------------------------------------|
| | Query Match | 17.0% | Score 242.5; | DB 1; | Length 451; |
| | Best Local Similarity | 27.4%; | Pred. No. 3.3e-11; | | |
| | Matches | 75; | Conservative | 49; | Mismatches 129; Indels 21; Gaps 7; |
| QY | 4 | VEAIGFCKFOWKLSVLTGLAWMADAMEMMTILSLAPOLHCEWRRLPSHQVALLTSVVVFVGM | 63 | | |
| Dd | 20 | LDRLPISRVHFQVLTAIGLYVFFDLADLFTISNVAPALIEHWGIPLTIANVTAAASFGLM | 79 | | |
| QY | 64 | MSSTLWGNISDQYGRKTGLKISVLWTLYCYLLSAFAPYSWIILVRGLVGFGIGGVPOS | 123 | | |
| Dd | 80 | FLGASLGGRUSDRIQRKKLANLFVVFETIASLCNAAMDIPSMLTFRFLTGFGVAANWI | 139 | | |
| QY | 124 | VTLTY-AEFLPMKARAKCILLIEVFWEAIGTFEVVLAVFVMP--SLGMRWLIIILSAYPLLL | 180 | | |

RT *Carnitine transporter OCTN2 mutations in systemic primary carnitine
 RT deficiency: a novel Arg169Gln mutation and a recurrent Arg282Ter
 RT mutation associated with an unconventional splicing abnormality. ;
 RL Biochem. Biophys. Res. Commun. 261:484-487(1999).
 RN [16]
 RP VARIANT CDSP CVS-211.
 RX MEDLINE-99408248; PubMed-10480371;
 RA Vaz F.M., Scholte H.R., Rulter J., Hussaarts-Odijk L.M.,
 RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,
 RA Wanders R.J.A.;
 RT *Identification of two novel mutations in OCTN2 of three patients with
 RT systemic carnitine deficiency. ;
 RL Hum. Genet. 105:157-161(1999).
 RN [17]
 RP VARIANT CDSP LEU-478.
 RX MEDLINE-99172075; PubMed-10072434;
 RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,
 RA Wanders R.J., Fox T.F., Helm N.M.;
 RT *Mutations of OCTN2, an organic cation/carnitine transporter, lead to
 RT deficient cellular carnitine uptake in primary carnitine deficiency. ;
 RL Hum. Mol. Genet. 8:655-660(1999).
 RN [18]
 RP CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.
 RX MEDLINE-20026865; PubMed-10559218;
 RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;
 RT *Mutations in novel organic cation transporter (OCTN2), an organic
 RT cation/carnitine transporter, with differential effects on the
 RT organic cation transport function and the carnitine transport
 RT function. ;
 RL J. Biol. Chem. 274:33388-33392(1999).
 RN [19]
 RP VARIANTS CDSP ARG-283 AND PHE-446.
 RX MEDLINE-20081068; PubMed-10612840;
 RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,
 RA Tsuji A.;
 RT *Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a
 RT patient with primary systemic carnitine deficiency. ;
 RL Hum. Mutat. 15:118-118(2000).
 RN [10]
 RP VARIANT CDSP LYS-452.
 RX MEDLINE-20145665; PubMed-10679939;
 RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;
 RT *A missense mutation in the OCTN2 gene associated with residual
 RT carnitine transport activity. ;
 RL Hum. Mutat. 15:238-245(2000).
 CC [1]
 CC FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC [1-] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC [1-] TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,
 CC HEART AND PLACENTA.
 CC [1-] DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY
 CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE
 CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE
 CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND
 CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL
 CC MYOPATHY OR CARDIOMYOPATHY.
 CC [1-] SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF057164; AAC24828.1; ;
 CC DR EMBL; AB015050; BAA29023.1; ;
 CC DR EMBL; AB016625; BAA36712.1; ;
 CC MIM; 603377; ;

MIM; 212140; ;
 DR InterPro: IPR003662; sub_transport.
 DR Pfam: PF00083; sugar_tr; 1.
 DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
 KW Transmembrane; Glycoprotein; Disease mutation.
 FT TRANSFEM 21
 FT TRANSFEM 41
 FT TRANSFEM 143
 FT TRANSFEM 163
 FT TRANSFEM 173
 FT TRANSFEM 193
 FT TRANSFEM 196
 FT TRANSFEM 216
 FT TRANSFEM 233
 FT TRANSFEM 253
 FT TRANSFEM 258
 FT TRANSFEM 278
 FT TRANSFEM 343
 FT TRANSFEM 363
 FT TRANSFEM 372
 FT TRANSFEM 392
 FT TRANSFEM 414
 FT TRANSFEM 434
 FT TRANSFEM 437
 FT TRANSFEM 457
 FT TRANSFEM 489
 FT CARBOHYD 57
 FT CARBOHYD 57
 FT CARBOHYD 64
 FT CARBOHYD 64
 FT CARBOHYD 91
 FT CARBOHYD 91
 FT VARIAT 169
 FT VARIAT 211
 FT VARIAT 211
 FT VARIAT 283
 FT VARIAT 283
 FT VARIAT 446
 FT VARIAT 446
 FT VARIAT 452
 FT VARIAT 452
 FT VARIAT 478
 FT VARIAT 478
 FT MUTAGEN 352
 FT MUTAGEN 352
 FT SEQUENCE 557 AA; 62751 MW; 928B1F6EFF63C48D CRC64;
 SQ
 Query Match 16.2%; Score 231; DB 1; Length 557;
 Best Local Similarity 28.2%; Pred. No. 2.9e-10;
 Matches 68; Conservative 45; Mismatches 112; Indels 16; Gaps 5;
 QY 44 EWRL---PSMQVALLTSVVFVGMSSSTLWGNISDQYGRKTGLKISVLVTLTYGILSAFA 100
 DB 131 EWNLCVDDMKAPLTISLFFVGVLLGSFISGQLSDRFGRKNVLFVTMGMTQSFSLQIFS 190
 QY 101 PVTYNILVRLGVFG-IGVPOSVTLYAEFLPKAKAKCLL-IEVFWAIGTVFEVLA 158
 DB 191 KNFEMFVVLVFLVGMGQISNYVAAVFLGTEILGKSVRIIFSTLGVCIYAFGYMVLPLFA 250
 QY 159 VFVMSLGNRWLLILSAVPLLLFAVLCLFVLPESARYDVLSGNOEKALATIKRIATENGAP 218
 DB 251 YFIR---DWRMLLVALTMPVCLVLMWFIPESPRWLISOGREFEEAEVIRKAAKANGIV 307
 QY 219 MP-----LGLIISROEDRGMKMDLFTPHPRWTTLTLLWFTWFSNAFSYGLVLLTTE 270
 DB 308 VPSTIEDPSELQDSSKKQSHNILDLLRTNIRMTVIMSIMLWMTISVGYFGLSLDTPN 367
 QY 271 L 271
 DB 368 L 368
 RESULT 9
 ID MUCKACICA STANDARD; PRT; 413 AA.
 AC P94131;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE CIS-CIS-MUCONATE TRANSPORT PROTEIN.
 GN MUCK
 OS Acinetobacter calcoaceticus.

```

OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OX Acinetobacter.
OX NCBI_TaxId=471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD413 / ADPI;
RX MEDLINE=97440147; PubMed=9294455;
RA Williams P.A., Shaw L.E.;
RT "muck, a gene in Acinetobacter calcoacetatus ADPI (BD413), encodes the
KT ability to grow on exogenous cis,cis-muconate as the sole carbon
RT source.";
RL J. Bacteriol. 179:5935-5942(1997).
RR
CC -1- FUNCTION: PROBABLE UPTAKE OF MUCONATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U87258; AAC27117.1; -.
CC InterPro; IPR003662; sub.transprtr.
CC Pfam; PF00083; sugar_tr; 1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC DR Transport; Transmembrane; Inner membrane.
CC KW TRANSMEM 17 37 POTENTIAL.
CC FT TRANSMEM 54 74 POTENTIAL.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 113 133 POTENTIAL.
CC FT TRANSMEM 146 166 POTENTIAL.
CC FT TRANSMEM 172 192 POTENTIAL.
CC FT TRANSMEM 229 249 POTENTIAL.
CC FT TRANSMEM 267 287 POTENTIAL.
CC FT TRANSMEM 294 314 POTENTIAL.
CC FT TRANSMEM 319 339 POTENTIAL.
CC FT TRANSMEM 362 382 POTENTIAL.
CC FT TRANSMEM 383 403 POTENTIAL.
CC SQ SEQUENCE 413 AA; 45245 MW; 2D88CE31C4C5CC65 CRC64;
CC -----
Query Match 15.9%; Score 227.5; DB 1; Length 413;
Best Local Similarity 25.8%; Pred. No. 4e-10;
Matches 68; Conservative 53; Mismatches 106; Indels 37; Gaps 9;
Qy 10 GKFWKLSVLVTG-LAWMADAMEMMILSILAPQLHCEWRLPSQVALLTSVVFVGMSSST 68
Db 11 GSHTWKIAFLFAFLALLVDGADLLMLLSYLSINSAEFLNSTVEAGMLGSFTLAGMAIGGI 70
Qy 69 LWNISDOYGRKTKLTSVLWTYYGILS-AFAPVYSKIL--VLRLGLVGFGIGGVPSV- 124
Db 71 FGMWACDFRGR---VRVWISILFTSLTGLGTQSFIOGVLFURFFASJGLGSLYACN 127
Qy 125 TLYAEFLPMKARAKCILLIEVFVAIGTVFVFYAVFVYMPSLGWRWLLILSAVPLLLFAVL 184
Db 128 TLMAEYPTKYRTVTLGTLAGWTGVIVATLLAGWLIPDHGRWRFVFAIIPVLMAVLM 187
Qy 185 CFWLPESARYDVLSGNOKAIAITLKRIATENGAMPWLGKLIISQEDRGKMRDLFTFHR 244
Db 188 HFFVPEPAW-----QOSRLAPSKQTETVKTSAF---KLIFQDKRRN----- 227
Qy 245 WTTLLWFIWFSNA----FSYYGL 264
Db 228 -----MEILWALTAGFLOFGYGV 246

```

[illegible]

```

Db 253 LFAEILCRPLLGVTLLMGVFF 274
::
|||||::|
RESULT 11
ID OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q920E8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).
GN SLC22A5 OR OCTN2.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99113835; PubMed=9916797;
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,
RA Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T.,
RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,
RA Tsuji A.;
RT "Primary systemic carnitine deficiency is caused by mutations in a
RT gene encoding sodium ion-dependent carnitine transporter.";
RL Nat. Genet. 21:91-94(1999).
[2]
RP SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.
RC STRAIN=C3H;
RX MEDLINE=99057546; PubMed=9837751;
RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;
RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine
RT cotransporter, in the juvenile visceral steatosis mouse.";
RL Biochem. Biophys. Res. Commun. 252:590-594(1998).
[3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter.";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL
CC STEATOSIS (JVS).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AB015800; BAA36590.1; -
DR EMBL; AF111425; AAC99787.1; -
DR EMBL; AF110417; AAD54060.1; -
DR MGD; MGI:1329012; Slc22a5.
DR InterPro; IPR003662; sub_transprtr.
DR Pfam; PF00083; sugar_tr; 1.
DR ProSite; PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein; Disease mutation.
FT TRANSMEM 21
POTENTIAL.

```

```

FT TRANSMEM 143 163
FT TRANSMEM 173 193
FT TRANSMEM 198 218
FT TRANSMEM 233 253
FT TRANSMEM 258 278
FT TRANSMEM 342 362
FT TRANSMEM 374 394
FT TRANSMEM 407 427
FT TRANSMEM 431 451
FT TRANSMEM 489 509
FT CARBOHYD 57 57
FT CARBOHYD 64 64
FT CARBOHYD 91 91
FT CARBOHYD 322 322
FT VARIANT 352 352
SQ SEQUENCE 557 AA; 62779 MW; 6093F0EE9612B204 CRC64;
Query Match 15.3%; Score 219; DB 1; Length 557;
Best Local Similarity 27.9%; Pred. No. 2.2e-09;
Matches 69; Conservative 43; Mismatches 107; Indels 28; Gaps 6;
QY 44 EHRL----PSQVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAPA 100
Db 131 EMDLVCKDDMKAPLTLSLFFVGLMGFSISGQLSDRGRKNVFLTLTMGMQTGFSFLOVFS 190
QY 101 PVSWTILVRLGLVGFQ-IGVQPSQVTLVYAEFLPMKARAKCILL-IEVFWAIGTVFEVFLA 158
Db 191 VNFEMFTVFLVGLMGQISNYVAAFLVGLTEILSKSRIIFATLCVGLCFYAFGFWLPLFA 250
QY 159 VFVPSLGRWRLLILSAVPLLLFAVLGFWLPESARYDVLSGNOEKAIAIUKRIATENGAP 218
Db 251 YFIR---DWRMLLLALTVPGLGALWFWFIPESPRWLISQGRKEAEVIRKAAKINGIV 307
QY 219 MPLGLKLIISROEDRGKMRDL-----FPHFRWTTLLLWFWFSNAFSYGL 264
Db 308 AP-----STIFDPSELQDLNKTSPOLHHYDILIRNIRVITIMSILLMLTISVGTFGL 361
QY 265 VLLTTEL 271
Db 362 SLDTPLN 368

```

```

RESULT 12
ID OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q70594; Q9QWLO;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)
DE (JVS2R) (CT1).
GN SLC22A5 OR OCTN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=98200080; PubMed=9541011;
RA Schoemig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,
RA Gruendemann D.;
RT "Molecular cloning and characterization of two novel transport
RT proteins from rat kidney.";
RL FEBS Lett. 425:79-86(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
RX MEDLINE=99011422; PubMed=9792817;
RA Sekine T., Kusahara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
RA Kanai Y., Endou H.;

```

RT Molecular cloning and characterization of high-affinity carnitine transporter from rat intestine.*;
RL Biochem. Biophys. Res. Commun. 251:586-591(1998).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter.*";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF CARNITINE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE BRAIN.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC CATION SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AJ001933; CAA05106.1; -;
DR EMBL; AB017260; BAA34399.1; -;
DR EMBL; AF110416; AAD54059.1; -;
DR InterPro: IPR003662; sub_transportr.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT TRANSMEM 489 509 POTENTIAL.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 275 275 W -> G (IN REF. 2).
SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;

Query Match 15.3%; Score 219; DB 1; Length 557;
Best Local Similarity 27.9%; Pred. No. 2.2e-09;
Matches 69; Conservative 43; Mismatches 107; Indels 28; Gaps 6;

QY 44 EMBL---PSWQVALLTSVVFVGMSSSTLWGNISDOYGRKTKLSVLWTLVYGLLSAFA 100
DB 131 EMDLVCKDDWRAPLTSFFGVLMGMSFISQLSDFGRKKNVLFITMGQTGFSFLQIFS 190
QY 101 PYSIWILRLGLVGRFG-IGGVPSQSVTLVYAEFLPMKARAKCILL-IEVFWAIGTVFEVILA 158
DB 191 VNFEMFTVLVLGVGQISNVVAAFLVTEILSKSIRIIFATLGVCIYVAFGFMVLPFLFA 250
QY 159 VFMVSLGWRWLLILSAVPLLLFAVLFCWLPESARYDVLGNOEKAIATLKRATENGAP 218
DB 251 YFIR---DWRMLLLALTVPGLCGALWWFIPESPRKLISQGRVKAEVIRKAAKENGIV 307

QY 219 MPLGKLLIISQEDRGKMRDL-----ETPHERWTLLLWFIFWNAESYVGL 264
DB 308 AP-----STIFDPSLODLNKKKQSHHIYDLVTRNIRIITIMSIILWLTISVGVFGL 361
QY 265 VLLTTEL 271
DB 362 SLDTFNL 368
RESULT 13
YGCs_ECOLI
ID YGCs_ECOLI STANDARD; PRT: 445 AA.
AC 045909;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YGCs.
GN YGCs OR B2771.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.*";
RL Science 277:1453-1474(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U29579; AAA69281.1; ALT_INIT.
DR EMBL; AE000360; AAC75813.1; ALT_INIT.
DR EcoGene: EGI3126; YGCs.
DR InterPro: IPR003662; sub_transportr.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 287 307 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 370 390 POTENTIAL.
FT TRANSMEM 401 421 POTENTIAL.
SQ SEQUENCE 445 AA; 48234 MW; B59E452721B15774 CRC64;

Query Match 14.4%; Score 205.5; DB 1; Length 445;
Best Local Similarity 25.3%; Pred. No. 1.8e-08;
Matches 72; Conservative 44; Mismatches 132; Indels 37; Gaps 7;

QY 4 VEAIGFGKFWKLVLTGLAWMADAMMMILSILAPQLHCEWRLPSQVALLTSVVFVGM 63
 Db 8 MDPLNRFHCRHIAALTEGHLTDGYVLGVIGYAIQLTPAMQLTFPMAGMIGGSSALLGL 67
 QY 64 MSSTLWGNISDYGKTKGLKISVLWLTLYGILSAFAPYISWILVRLGLVFGIGGVQPS 123
 Db 68 FLGSLVLGWSIDHIGRQKTFESFLILITLASFLOFFATTPEHLILIGLIGLGGC-DYS 126
 QY 124 V-TLYAEFLPMKAKACILLIEVFWAIGTVFVVLAVFVMPSL-----GWRWL 170
 Db 127 VGHUTLAEESPRRHGILGAFSVVTVG-----YVLASIGHFISENPENRWL 177
 QY 171 LILSAVPLLLFAVLCFVLPESARYDVLGNQEKATATKRIATENGAPMLGKLILISROE 230
 Db 178 LASAALPALLITLRLWGTPTSPRWLLRQGRFAEHAHVHYE---GPHVLLGDVVVTAH 234
 QY 231 DRCK-----MRDLTFPHFWTLLWFI---WFSNAFSYYGL 264
 Db 235 KHIKTLFSSRYRRRTAFNSVFFVCLVFPWFVIYTWLPTIAQTIGL 279

RESULT 14
 MMLH_ALCEU STANDARD; PRT; 428 AA.
 AC 051798;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PROBABLE 4-METHYLMUCONOLACTONE TRANSPORTER.
 GN MMLH;
 OS Alcaligenes eutrophus (Ralstonia eutropha).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JMP134;
 RX MEDLINE=98121313; PubMed=9461415;
 RA Erb R.W., Timmis K.N., Pieper D.H.;
 RT "Characterization of a gene cluster from Ralstonia eutropha JMP134 encoding metabolism of 4-methylmucunolactone.";
 RL Gene 206:53-62(1998).
 CC -!- FUNCTION: PROBABLE UPTAKE OF 4-METHYLMUCONOLACTONE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; X99639; CAA67957.1; -
 DR InterPro: IPR003662; sub.trnsportr.
 DR Pfam: PF000083; sugar.tr.1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 KW Transport; Transmembrane.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 146 166 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 253 273 POTENTIAL.
 FT TRANSMEM 288 308 POTENTIAL.
 FT TRANSMEM 315 335 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 FT TRANSMEM 378 398 POTENTIAL.
 SQ SEQUENCE 428 AA; 45804 MW; 7CC9096F6F7BE230 CRC64;

Query Match 14.2%; Score 203.5; DB 1; Length 428;
 Best Local Similarity 27.4%; Pred. No. 2.5e-08;
 Matches 49; Conservative 38; Mismatches 83; Indels 9; Gaps 4;
 QY 24 WNADEMWMILSILAPQLHCEWRLPSQVALLTSVVFVGMSSSTLWGN--ISDQYGRKT 81
 Db 23 WALDSFDMQMSFLLPALTLTWGLTKAEYGVLTVALV--VTAIGWAGAGILSDRYGRAR 80
 QY 82 GLKISVLWLTLYGILSAFAPYISWILVRLGLVFGIGGV-POSVTLYAEFLPMKAKACI 140
 Db 81 ILVLAIIWTLTGLVLAGFAQSTQQLLIARTLQGLGGEWAYCAALMAEVIDSRHKGAI 140
 QY 141 LLIE----VFWAIGTVFVVLAVFVMPSLGWRMLLILSAVPLLLFAVLCFVLPESARYD 195
 Db 141 GFVQSGFALGWALVVVATLLLAWLPKEMARVAFWSGIIPALIVLFIIRHVKSSMFE 199

RESULT 15
 STPL_ARATH STANDARD; PRT; 522 AA.
 AC P23586;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE GLUCOSE TRANSPORTER (SUGAR CARRIER).
 GN STPL.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERRECTA;
 RX MEDLINE=91005995; PubMed=2209537;
 RA Sauer N., Friedlaender K., Graeme-Wicke U.;
 RT "Primary structure, genomic organization and heterologous expression of a glucose transporter from Arabidopsis thaliana.";
 RL EMBO J. 9:3045-3050(1990).
 CC -!- FUNCTION: ACTIVE UPTAKE OF HEXOSES. PROBABLE GLUCOSE/HYDROGEN SYMPORT.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; X55350; CAA39037.1; -
 DR PIR; S12042; S12042.
 DR InterPro: IPR003663; Sugar_trnsportr.
 DR InterPro: IPR003662; sub.trnsportr.
 DR Pfam: PF00083; sugar.tr.1.
 DR PRINTS; PR00171; SUGTRNSPORT.
 DR PRINTS; PR00172; GLUCTRNSPORT.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 KW Transmembrane; Transport; Sugar transport; Symport.
 FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 22 42 POTENTIAL.
 FT TRANSMEM 84 104 POTENTIAL.
 FT TRANSMEM 113 133 POTENTIAL.
 FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 285 305 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.

```

FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT DOMAIN 474 522 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 522 AA; 57596 MW; 68A6C72AFFD90380 CRC64;

Query Match 14.0%; Score 200; DB 1; Length 522;
Best Local Similarity 24.0%; Pred. No. 5.3e-08;
Matches 69; Conservative 59; Mismatches 135; Indels 24; Gaps 9;

QY 4 VEAIGFGKFWKLSVLTGLAWMADAMEMMILSILAPQ-----LHCEWRLPNQVALLT 56
   | | | | | : : : | : : | : | : | : | : | : | : | : | : | : |
Db 29 VAAMGGLIFGYDIGSGVTSMPSPFLARFFPSVYRKQOEDASTNQYCYDSDPT--LTMFT 86
   | : : : | | : : : | : : | : : | : : | : : | : : | : : | : : |
QY 57 SVFVGMSSSTLWGNISDQYGRKTGLKISVLWTLTYGILSAPAPYYSWILVL-RGLVGF 115
   | : : : | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 87 SSLYLAALISLVASTVTRKFGRLSMLFGGILFCAGALINGFAK-HVWMLIVGRILLGF 145
   | | | | | : | | | | | : : : | : : | : : | : : | : : | : : |
QY 116 GIGGVPSQVTLY-AEFLPMKARAKCILLIEVFAIG-TVEVVLAVFVMPSLGWRLLIL 173
   | | | | | : | | | | | : : : | : : | : : | : : | : : | : : |
Db 146 GIGFANQAVPLYLSEMAPYKYRGALNIGFQLSITIGILVAEVLNFFAKIKGCGWRLSL 205
   | | | | | : | | | | | : : : | : : | : : | : : | : : | : : |
QY 174 --SAVPLLLFAVLCPWPESARYDVLSGNOEKAATLKRATENGAMPPLGKLIISQED 231
   | | | : | : : | : : | : | : | : | : | : | : | : | : | : |
Db 206 GGAVVPALIIITIGSLVLPDTPNSMIERQHEEAKTKLRRIRGVDDVSQOEFDDLVAASKES 265
   | | | : | : : | : : | : | : | : | : | : | : | : | : | : |
QY 232 -----RGKMRDLTFPHFRWTTLLWFIFS--NAFSYYGLVLLTT 269
   | | | : | : : | : : | : | : | : | : | : | : | : | : | : |
Db 266 QSIHPWRNLLRRKRYRPHLTMAVMIPFFQQLTGINVMFYAPVLFNT 312
   | | | : | : : | : : | : | : | : | : | : | : | : | : | : |
```

Search completed: March 13, 2002, 12:42:53
Job time: 159 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:42:29 ; Search time 43.34 Seconds
(without alignments)
921.374 Million cell updates/sec

Title: US-09-911-667A-4
Perfect score: 1429
Sequence: 1 EDVAEAIQFGKFWKLSVLT.....WFSNAPSYYGLVLLTTELPQ 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 1409 | 98.6 | 548 | 11 Q92217 | Q92217 rattus norv |
| 2 | 751.5 | 52.6 | 497 | 5 Q9W1D4 | Q9W1D4 drosophila |
| 3 | 617.5 | 43.2 | 470 | 10 Q9LSH7 | Q9LSH7 arabidopsis |
| 4 | 609.5 | 42.7 | 500 | 10 Q9M7W2 | Q9M7W2 arabidopsis |
| 5 | 446.5 | 31.2 | 454 | 2 Q9RYN9 | Q9RYN9 deinococcus |
| 6 | 358.5 | 25.1 | 422 | 2 Q9X1T9 | Q9X1T9 thermotoga |
| 7 | 344 | 24.1 | 727 | 11 Q9Z216 | Q9Z216 rattus norv |
| 8 | 342 | 23.9 | 742 | 11 Q9J1S5 | Q9J1S5 mus musculu |
| 9 | 341.5 | 23.9 | 709 | 5 Q9W3W9 | Q9W3W9 drosophila |
| 10 | 341 | 23.9 | 724 | 13 Q9Q4O6 | Q9Q4O6 discopyge o |
| 11 | 338 | 23.7 | 607 | 4 Q9BVZ9 | Q9BVZ9 homo sapien |
| 12 | 338 | 23.7 | 742 | 4 Q94841 | Q94841 homo sapien |
| 13 | 338 | 23.7 | 742 | 6 Q9397 | Q9397 bos taurus |
| 14 | 331 | 23.2 | 683 | 4 Q94840 | Q94840 homo sapien |
| 15 | 320 | 22.4 | 455 | 2 Q9HYD9 | Q9HYD9 pseudomonas |
| 16 | 320 | 22.4 | 683 | 11 Q63564 | Q63564 rattus norv |
| 17 | 298 | 20.9 | 443 | 2 Q9L7Y1 | Q9L7Y1 pseudomonas |
| 18 | 281 | 19.7 | 556 | 11 Q63089 | Q63089 rattus norv |
| 19 | 281 | 19.7 | 593 | 11 P70485 | P70485 rattus norv |

```

20      280      19.6      555      11      Q9BOW2      Q9BOW2 rattus norv
21      280      19.6      593      11      P97558      P97558 rattus norv
22      278      19.5      536      11      Q9RIU7      Q9RIU7 rattus norv
23      277      19.4      553      11      O70577      O70577 mus musculu
24      275.5     19.3      556      11      O08966      O08966 mus musculu
25      275      19.2      554      6      O02713     O02713 sus scrofa
26      275      19.2      556      11      Q9RIQ4      Q9RIQ4 mus musculu
27      273      19.1      555      4      O15244     O15244 homo sapien
28      271.5     19.0      535      11      Q63314     Q63314 rattus norv
29      270.5     18.9      537      11      O88909     O88909 mus musculu
30      267.5     18.7      448      2      Q9I6Q3     Q9I6Q3 pseudomonas
31      267.5     18.7      562      13      O57379     O57379 pseudopleur
32      265      18.5      446      2      Q9HUR3     Q9HUR3 pseudomonas
33      264      18.5      448      2      Q9I110     Q9I110 pseudomonas
34      264      18.5      526      10      Q9LHQ6     Q9LHQ6 arabidopsis
35      258      18.1      465      2      Q9AQ12     Q9AQ12 pseudomonas
36      257      18.0      548      5      O01384     O01384 drosophila
37      257      18.0      548      5      Q9VCA2     Q9VCA2 drosophila
38      257      18.0      554      4      O15395     O15395 homo sapien
39      255      17.8      554      4      O15245     O15245 homo sapien
40      251      17.6      554      6      Q77504     Q77504 oryctolagus
41      249      17.4      551      4      Q9Y226     Q9Y226 homo sapien
42      248.5     17.4      463      2      Q9RHQ1     Q9RHQ1 variororax
43      247      17.3      554      4      Q9NQD4     Q9NQD4 homo sapien
44      246      17.2      567      5      Q9VCA3     Q9VCA3 drosophila
45      245      17.1      430      11      O35882     O35882 rattus norv
```

ALIGNMENTS

```

RESULT 1
Q92217 ID Q92217 PRELIMINARY; PRT; 548 AA.
AC Q92217;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SV2 RELATED PROTEIN.
GN SVOP.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=99019745; PubMed=9801366;
RA Janz R., Hofmann K., Sudhof T.C.;
RT "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests
RT novel transport functions of synaptic vesicles.";
RL J. Neurosci. 18:9269-9281(1998).
CC -!- SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF060173; AAC78627.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugaf_tr; 1.
KW Transmembrane.
SQ SEQUENCE 548 AA; 60804 MW; 19AD8475B7579496 CRC64;
```

Query Match 98.6%; Score 1409; DB 11; Length 548;
Best Local Similarity 97.4%; Pred. No. 1.3e-91;
Matches 266; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 EDVAEAIQFGKFWKLSVLTGLAWMADAMEMILSLAPLHCEWRLPSQVALLTSVVF 60
      |||||
DB 71 EDVAEAIQFGKFWKLSVLTGLAWMADAMEMILSLAPLHCEWRLPSQVALLTSVVF 130
      |||||
QY 61 VGMSSSTLWGNISDQYGRKTKGLKISVLTWYTGILSAFAPYVSWTLVLRLGVGFGIGV 120
      |||||
DB 131 IGMSSSTLWGNISDQYGRKTKGLKISVLTWYTGILSAFAPYVSWTLVLRLGVGFGIGV 190
```


QY 121 POSVTLYAEFLPMKARAKOILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLL 180
DB 191 POSVTLYAEFLPMKARAKOILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLV 250
QY 181 FAVLCFWLPESARYDVLGSGQEKATATLKRIATENGAPMLGKLIISROEDRGKWRDLFT 240
DB 251 FAVLCFWLPESARYDVLGSGQEKATATLKRIATENGAPMLGKLIISROEDRGKWRDLFT 310
QY 241 PFRWTTLLWFIWFSNAFSYYGLVLLTTFELFQ 273
DB 311 PFRWTTLLWFIWFSNAFSYYGLVLLTTFELFQ 343
RESULT 2
Q9W1D4 PRELIMINARY: PRT: 497 AA.
AC Q9W1D4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG4324 PROTEIN.
GN CG4324.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazie R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.B.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Nusskern D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC CR -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DB EMBL; AE003462; AAF47135.1; -;
DB FlyBase; FBgn0034956; CG4324.

DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 497 AA; 55360 MW; 678F4A3286B7BD4 CRC64;
Query Match 52.6%; Score 751.5; DB 5; Length 497;
Best Local Similarity 53.7%; Pred. No. 2e-45;
Matches 146; Conservative 46; Mismatches 79; Indels 1; Gaps 1;
QY 1 EDVAEIAFGKFGKFWKLSVLTGLAWMADAMEMMILSILAPOLHCENRPLSPQVALLTSVVF 60
DB 62 QQAIAAFGFGFHVKLSLLVGLGWNDSMEMAILSLGSLFCEWNVTKFQQAASVTTVVF 121
QY 61 VGMSSSTLWGNISDOYGRKTKGLKISVLTWLTLYYGILSAFAPVYSWILVIRGLVGIGCV 120
DB 122 LGMMLSSFTWQLSNRYGRKSAITLFGVLLVLYLSLVSVAPSYAWLLTLRGLVGAIGCV 181
QY 121 POSVTLYAEFLPMKARAKOILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLL 180
DB 182 POSVTLYAEFLPMKARAKOILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLL 241
QY 181 FAVLCFWLPESARYDVLGSGQEKATATLKRIATENGAPMLGKLIISROEDRGK-WROLF 239
DB 242 FTILSPWSESARYSYNGHNDKAIKVLQIAHNKRHMLMGLMADDEPSCAESFRSL 301
QY 240 TPFHRTWTTLLWFIWFSNAFSYYGLVLLTTFEL 271
DB 302 SPSLYRTTLLWFLWLASAFYCYGLVLTTEL 333
RESULT 3
Q9LSH7 PRELIMINARY: PRT: 470 AA.
AC Q9LSH7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRANSPORTER-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty pl and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DB EMBL; AB026645; BAB02515.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 470 AA; 51457 MW; 1062413BA354616D CRC64;
Query Match 43.2%; Score 617.5; DB 10; Length 470;
Best Local Similarity 43.5%; Pred. No. 5e-36;
Matches 121; Conservative 52; Mismatches 98; Indels 7; Gaps 1;
QY 1 EDVAEIAFGKFGKFWKLSVLTGLAWMADAMEMMILSILAPOLHCENRPLSPQVALLTSVVF 60

Db 11 DEALVAMGCKFKQIYVLAYAGMGWAEAMEMMLLSFVGPAVQSLWNLNSARQESLITSVF 70
Qy 61 VGMSSSTLWGNISDOYGRKTKLISVLTYLILYGLSAPAFYVSWILVRLGLVGFGGV 120
Db 71 AGMLIGAYSWGIVSKHGRKGFIIITAVVTFVAGFSAFSPNTYMWLLIILRLCLVGLGCGG 130
Qy 121 POSVTLYAEFLPMKARAKCILLIEVFVWAGTVEFVLAFFVMPSLGWRWLLIILSAVPLLL 180
Db 131 PVLASWYLEFIPAPSRCTWNVVFAFWTGTIFEASLAWLMPRLGWRWLLAFSSVPSL 190
Qy 181 FAVLCFWLPESARYDVLSGNOEKATLKRATENGAPMPLGK-----IISRQDRG 233
Db 191 LLLFYRWTSFPRYLILQGRKAELALEKRIARNKTKQLPPGVLSSELETELENKEPGF 250
Qy 234 KMRDLFTPHFRWTTLLWFWFNSAFSYGLVLLTTTEL 271
Db 251 SLLALLSPTLMKRTLLLWVFFGNFAYYGVLLTTTEL 288
RESULT 4
Q9M7W2 ID Q9M7W2 PRELIMINARY; PRT; 500 AA.
AC Q9M7W2; 2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE PUTATIVE TRANSPORTER.
GN MGH6.16.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III P1 MGH6 genomic sequence".
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AC024128; AAF35954.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
DR Transmembrane.
KW SEQUENCE 500 AA; 54573 MW; EC459EB09EF581C5 CRC64;
Query Match 42.7%; Score 609.5; DB 10; Length 500;
Best Local Similarity 39.3%; Pred. No. 2e-35;
Matches 121; Conservative 55; Mismatches 95; Indels 37; Gaps 2;
Qy 1 EDVAIGFGKFGKQKLSVLTGLANMADAMEMMILSILAPQLHCEWRLPSQVALLTSVVF 60
Db 11 DEALVAMGFGFKQIYVLAYAGMGWAEAMEMMLLSFVGPAVQSLWNLNSARQESLITSVF 70
Qy 61 VGMSSSTLWGNISDOYGRKTKLISVLTYLILYGLSAPAFYVSWILVRLGLVGFGGV 120
Db 71 AGMLIGAYSWGIVSKHGRKGFIIITAVVTFVAGFSAFSPNTYMWLLIILRLCLVGLGCGG 130
Qy 121 POSVTLYAEFLPMKARAKCILLIEVFVWAGTVEFVLAFFVMPSLGWRWLLIILSAVPLLL 180
Db 131 PVLASWYLEFIPAPSRCTWNVVFAFWTGTIFEASLAWLMPRLGWRWLLAFSSVPSL 190
Qy 181 FAVLCFWLPESARYDVLSGNOEKATLKRATENGAPMPLG-----IISRQDRG 233
Db 191 LLLFYRWTSFPRYLILQGRKAELALEKRIARNKTKQLPPGVLSSELETELENKNIPT 250
Qy 223 -----KLIISRQDRG-KMRDLFTPHFRWTTLLWFWFNSAFSYG 263

Db 251 ENTHLLKAGSGEAVAVSKIVLAKRDEPGFSLALLSPTLMKRTLLWVFFGNFAYY 310
Qy 264 LVLLTTTEL 271
Db 311 VVLLTTTEL 318
RESULT 5
Q9RYN9 ID Q9RYN9 PRELIMINARY; PRT; 454 AA.
AC Q9RYN9;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE SUGAR TRANSPORTER, PUTATIVE.
GN DRA0271.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RC MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C., D.,
RA Makareva K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI".
RL Science 286:1571-1577(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE001863; AAF12486.1; -.
DR TIGR; DRA0271; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Complete proteome; Sugar transport; Transmembrane.
KW SEQUENCE 454 AA; 48171 MW; 214EA1A3EDC60B8B CRC64;
Query Match 31.2%; Score 446.5; DB 2; Length 454;
Best Local Similarity 35.1%; Pred. No. 5.1e-24;
Matches 94; Conservative 52; Mismatches 115; Indels 7; Gaps 3;
Qy 4 VEAIGFGKFGKQKLSVLTGLANMADAMEMMILSILAPQLHCEWRLPSQVAVV 61
Db 19 LDDLGGRFGKQKLSVLTGLANMADAMEMMILSILAPQLHCEWRLPSQVAVV 78
Qy 62 GMMSSSTLWGNISDOYGRKTKLISVLTYLILYGLSAPAFYVSWILVRLGLVGFGGV 120
Db 79 GMLFGAMFGVGLADRVGRSRVFLTVAGVVGFGALAPLTLTWLLVAFRLTGAIGTLL 138
Qy 121 POSVTLYAEFLPMKARAKCILLIEVFVWAGTVEFVLAFFVMPSLGWRWLLIILSAV 176
Db 139 PVDYSMAAEFVPTAWRGFVLYLESFVNAVGVVVAALAWWSTAFAPAEGRWLLGLAAL 198
Qy 177 PLLLFVLCFWLPESARYDVLSGNOEKATLKRATENGAPMPLGKLIISRQDRGKMR 236
Db 199 PGLVGLIARIIGIPDSRSLARGEAEQARAALQKVAQANGGTLPAPLAHPQEPFRVSPA 258
Qy 237 DLFTPHFRWTTLLWFWFNSAFSYG 264
Db 259 QLFRGVLAARRPTLLWFWFNSAFSYG 286
RESULT 6
Q9X1T9 ID Q9X1T9 PRELIMINARY; PRT; 422 AA.

```
AC O9XLT9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PERMEASE, PUTATIVE.
GN TM1603.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AE001804; AAD36670.1; -.
DR TIGR; TM1603; -.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00083; sugat_tr; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBER; UNKNOWN_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Complete proteome; Transmembrane.
SQ SEQUENCE 422 AA; 47152 MW; 13f9DC5649A1338D CRC64;

Query Match 25.1%; Score 358.5; DB 2; Length 422;
Best Local Similarity 32.0%; Pred. No. 7.4e-18;
Matches 89; Conservative 55; Mismatches 111; Indels 23; Gaps 7;

QY 1 EDVAEAGFGKQKLSVLTGLAWMADAMEMMILSILAPQLHCWRLPSQVALLTSVVF 60
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 4 DEIVEKYVDKTRQRFILTSIAWFDAAAGVLLSFVLPIYIKENLNTSTQCATIASATF 63

QY 61 VGMSSSTLWGNISDQYGRKTKLSVLWTLYYGILSAFAPYVSWILVRLGLVFGIGGV 120
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 64 LGMLFGALSVGEVDLLGRKVSNNLFFVITITFTLSGFSFSETLLVRLGLSGFGYGL 123

QY 121 PQSVTLV-AEFLPMKARAKCILLIEVFWAIGTVFEVILAVFVMPSLGWRLLILSAVPLL 179
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 124 MFSFNAYLAETFSIRLGRYLVLLESSWAVGSILIGLFAVNVLPN--WRWVFWIFSIG-Y 180

QY 180 LFAVLCFWLPESARYDVLSGNOKAIAIYL-KRIATENGAP-----MPLGKLIISROEDRGK 234
   | | :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 181 LFPVFLRMPETPKYAFLLGKGEALERSLGRVVEEVLPELKEKYPILALL-----KREH 235

QY 235 MRDLTPHFRWTLTLWIFSNAPSYGYGLVLLTTELF 272
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 236 LKD-----TVVILWIAFWVSVYYALFTWAPRIF 264

RESULT 7
Q92216 PRELIMINARY; PRT; 727 AA.
ID Q92216
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE SYNAPTIC VESICLE PROTEIN 2C.
GN SV2C.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99019745; PubMed=9801366;
RA Janz R., Hofmann K., Sudhof T.C.;
RT "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests
RT novel transport functions of synaptic vesicles.";
RL J. Neurosci. 18:9269-9281(1998).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF060174; AAC78628.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugat_tr; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 727 AA; 82248 MW; 4C154C69341D8DB7 CRC64;

Query Match 24.1%; Score 344; DB 11; Length 727;
Best Local Similarity 28.4%; Pred. No. 1.3e-16;
Matches 88; Conservative 57; Mismatches 115; Indels 50; Gaps 7;

QY 4 VEAGFGKQKLSVLTGLAWMADAMEMMILSILAPQLHCWRLPSQVALLTSVVFVGM 63
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 142 IOECGHRFPQWALFFVLGMALMADGVEFVGVFLPSAETDLCIPNSGSGWLSIVYLG 201

QY 64 MSSSTLWGNISDQYGRKTKLSVLWTLYYGILSAFAPYVSWILVRLGLVFGIGGV 122
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 202 MVGAFFWGGGLADKVGKQSLICMSVNGVFFAFLSFVQGYGFLLCRLLSGFGIGAIPT 261

QY 123 SVTLVLAELPMKARAKCILLIEVFWAIGTVFEVILAVFVMPSLG-----WRW 169
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 262 VFSYFAEVLAREKRGHLSWLCMFWMIGGIYASAMAWAIIPIHYGWSFGMSGAYQFHSRW 321

QY 170 LLILSAVPLLLPAVLFCFWLPESARYDVLSGNOKAIAIYL-KRIATEN-----MPLG 222
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 322 FVIVCALPCVSVVALTFMPESPRLLELVGKHDEAMMLKLHIDTNMRARGQPEKVFVN 381

QY 223 KL-----ILSROEDRGK-----MRDLTPHFRWTLTLWIFW 254
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 382 KIKTPQKIDELIEIESDTGTWYRRCFVRIRTELYGIWLTFR-CFNPVPRENTIKLTIV 440

QY 255 FSNAPSYGYGL 264
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 441 FTLSFGYGL 450

RESULT 8
Q9JIS5 PRELIMINARY; PRT; 742 AA.
ID Q9JIS5
AC Q9JIS5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CA2+ REGULATOR SV2A.
GN SV2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=2008300; PubMed=10624962;
RA Janz R., Goda Y., Geppert M., Missler M., Sudhof T.C.;
RT "SV2A and SV2B function as redundant Ca2+ regulators in
RT neurotransmitter release.";
RL Neuron 24:1003-1016(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF196781; AAF87321.1; -.
DR EMBL; AF196780; AAF87321.1; JOINED.
DR MGD; MGI:1927139; SV2a.
```

```

DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 742 AA; 82647 MW; 1074857FD13ED894 CRC64;

Query Match      23.9%; Score 342; DB 11; Length 742;
Best Local Similarity 27.4%; Pred. No. 1.9e-16;
Matches 86; Conservative 55; Mismatches 125; Indels 48; Gaps 5;

Qy 1 EDVAEAIQFGKFWKLSVLTGLAWADAMEMMILSLAPOLHCEWRLPSQWVALLTSVVF 60
Db 153 ETILRECHGGEFTWLTFLVGLALMADGVEFVVGFLPSAEKDMCLSDSKNGMLGLIVY 212
Qy 61 VGMSSSTLWGNISDOYGRKTKGLISWLTLYGILSAFAPVSWILVRLGLVFGIGG- 119
Db 213 LGMVGAFLWGLADRLGRQCLLSISVNSVFAFFSFVQGYCTFLFCRLLSGVGIGGS 272
Qy 120 VPOSVTLAEFLPMKARAKCILLLEVEFWAIGTVFEVVLAVFVMPSLG----- 166
Db 273 IPVFYSFEFLAEQKKEHLSWLCFMWVGCVYAAAMAWAIIIPHYGSFOMG;SAYQPHS 332
Qy 167 WRWLILSAVPLLLFAVLCEFWLPESARYDVLSGNOEKAIATLTKRIATEN-----M 219
Db 333 WRVFLVCAFFSVEAIGALTTPQSPRFFLENGKHDEAWMLVKOVHDTNMRKAGHPERVF 392
Qy 220 PLGKLIIISROEDR-----CKMRDLETPHFRWTLTLWLF 252
Db 393 SVTHIKTIHQEDELIEIQTSDGTGTWQWGRVRLSLGGVGVNGFLSCFSPFYRRITLMMWG 452
Qy 253 IWFNSAFSYGYGLV 266
Db 453 VNETSFSYGYGLTV 466

RESULT 9
Q9W3W9 PRELIMINARY; PRT; 709 AA.
ID AC Q9W3W9
RA Adamantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Iandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siding-Klamis I., Simpson M., Skupski M.P., Smith F.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE003438; AAF46193.1; -.
DR FlyBase; FBgn0029896; CG3168.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane.
KW Transmembrane.
SQ SEQUENCE 709 AA; 77930 MW; 1B5AD1E9D133AE94 CRC64;

Query Match      23.9%; Score 341.5; DB 5; Length 709;
Best Local Similarity 27.1%; Pred. No. 2e-16;
Matches 83; Conservative 65; Mismatches 123; Indels 35; Gaps 6;

Qy 1 EDVAEAIQFGKFWKLSVLTGLAWADAMEMMILSLAPOLHCEWRLPSQWVALLTSVVF 60
Db 215 ERAILLCGYKPHY ILLAICGLVSTSEMDVISMFLPSAECDLDTNTETKGLNSIIF 274
Qy 61 VGMSSSTLWGNISDOYGRKTKGLISVLTLYGILSAFAPVSWILVRLGLVFGIGGV 120
Db 275 IGMVGAFFWGSIAISFGKVKLVIVISFNNAFCIVASSFSQTSYFFMLFRFLNGAALGS 334
Qy 121 -POSVTLAEFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVMP-SLG-----WR 168
Db 335 GPVWSYFAEFQPKAKGSMLSFMAAFWTEGMLFVSLAWLIIPRTIGTTPYFTYNSWR 304
Qy 169 WLLILSAVPLLLFAVLCEFWLPESARYDVLSGNOEKAIATLTKRIATENGAPMPLGLIISR 221
Db 395 IFLVCSLPSFLVGLFLFLYLPESPKFLTRGKKDRALALFRGIFVNTKRRDEYVYDL 454
Qy 229 Q-----EDRGMKRDLF-----TPHFRWTLTLWLFNSAFSYGYGLV 266
Db 455 EVDEKLLSNGVKNKYSRMISGMVDHSRALFKSPILRFTIVSI-TINFTHIGYGYGLIM 513
Qy 267 LTTELF 272
Db 514 WPELF 519

RESULT 10
Q90406 PRELIMINARY; PRT; 724 AA.
ID AC Q90406
RA 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRANSMEMBRANE TRANSPORTER.
OS Discopyge ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hypnosquala; Pristioraja; Batoidea;
OC Torpediniformes; Narcinoidel; Narcinidae; Discopyge.
OX NCBI_Taxid=7785;

```

RESULT 12

RI for large proteins in vitro. ;
 RL Data Res. 5:277-286 (1998).
 -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 CC EMBL; AB018279; BAA34456.1; .
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
 KW Transmembrane.
 SQ SEQUENCE 742 AA; 82695 MW; 913E216D5CFC2FB2 CRC64;

Query Match 23.7%; Score 338; DB 4; Length 742;
Best Local Similarity 27.4%; Pred. No. 3.6e-16;
Matches 86; Conservative 54; Mismatches 126; Indels 48; Gaps 5;

| | | | |
|------|-----|--|-----|
| Qy | 120 | VQSVTLTAYEFLPMKRAKACILLTEVFWAICTVFVVLAVFVPSLG----- | 166 |
| D6 | | | |
| D7 | | | |
| D8 | | | |
| D9 | | | |
| D10 | | | |
| D11 | | | |
| D12 | | | |
| D13 | | | |
| D14 | | | |
| D15 | | | |
| D16 | | | |
| D17 | | | |
| D18 | | | |
| D19 | | | |
| D20 | | | |
| D21 | | | |
| D22 | | | |
| D23 | | | |
| D24 | | | |
| D25 | | | |
| D26 | | | |
| D27 | | | |
| D28 | | | |
| D29 | | | |
| D30 | | | |
| D31 | | | |
| D32 | | | |
| D33 | | | |
| D34 | | | |
| D35 | | | |
| D36 | | | |
| D37 | | | |
| D38 | | | |
| D39 | | | |
| D40 | | | |
| D41 | | | |
| D42 | | | |
| D43 | | | |
| D44 | | | |
| D45 | | | |
| D46 | | | |
| D47 | | | |
| D48 | | | |
| D49 | | | |
| D50 | | | |
| D51 | | | |
| D52 | | | |
| D53 | | | |
| D54 | | | |
| D55 | | | |
| D56 | | | |
| D57 | | | |
| D58 | | | |
| D59 | | | |
| D60 | | | |
| D61 | | | |
| D62 | | | |
| D63 | | | |
| D64 | | | |
| D65 | | | |
| D66 | | | |
| D67 | | | |
| D68 | | | |
| D69 | | | |
| D70 | | | |
| D71 | | | |
| D72 | | | |
| D73 | | | |
| D74 | | | |
| D75 | | | |
| D76 | | | |
| D77 | | | |
| D78 | | | |
| D79 | | | |
| D80 | | | |
| D81 | | | |
| D82 | | | |
| D83 | | | |
| D84 | | | |
| D85 | | | |
| D86 | | | |
| D87 | | | |
| D88 | | | |
| D89 | | | |
| D90 | | | |
| D91 | | | |
| D92 | | | |
| D93 | | | |
| D94 | | | |
| D95 | | | |
| D96 | | | |
| D97 | | | |
| D98 | | | |
| D99 | | | |
| D100 | | | |

QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNOEKAIAIATLKRIATEN-----GAP---M 219
ID 094840 PRELIMINARY; PRT; 683 AA.
AC 094840;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P87=TRANSPORTER-LIKE PROTEIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RQ MEDLINE=93050176; PubMed=1426240;
RA Gingrich J.A., Andersen P.H., Tiberi M., el Mestikawy S.,
RA Jorgensen P.N., Fremieu R.T., Jr., Caron M.G.;
RT "Identification, characterization, and molecular cloning of a novel
RT transporter-like protein localized to the central nervous system."
RL FEBS Lett. 312:115-122(1992).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: A47919; AAB24028.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 742 AA; 82589 MW; 58AE0C9BB224DEB8 CRC64;

Query Match 23.7%; Score 338; DB 6; Length 742;
Best Local Similarity 27.4%; Pred. No. 3.6e-16;
Matches 86; Conservative 54; Mismatches 126; Indels 48; Gaps 5;

QY 1 EDAYEATIGFGKFWKLSVLTGLAMMADAMEMMILSILAPQLHCEWRLPSWOVALLTSVVF 60
ID 153 EAILRECGHGRFQWTLFVLGLALMADGVEFVVGFLVPSAEKDMCLSDSNKMGILVIV 212
QY 61 VGMSSSTLWGNISDOYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG- 119
ID 213 LGMMVGAFLWGLADRLGRQCLLSLVNSVFAFFSSVQGYGTFGLFCRLLSGVGIGGS 272
QY 120 VPOSVTLVAEFLPMKARAKCILLIEVFVAIGTVFVWLVAVFVMPSLG----- 166
ID 273 IPIVSYSEFLAQEKREHLSWLCFMWIGGYAAAMAWAIIPHYGWSFQMGSAQPHS 332
QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNOEKAIAIATLKRIATEN-----GAP---M 219
ID 333 WRVFLVCAFPVFAIGALTTPQESPRFFLENGKHDEAWMLKQVHDTNMRAGHPERV 392
QY 220 PLGLKLIISRODR-----GKMRDLTFPHFRWTTLLWF 252
ID 393 SVTHKTHQDELEIEIOSDTGTQWQVGRVRLSLGGQVNGFLSCFGPEYRITILMMMG 452
QY 253 IWFSAFYSYGLVL 266
ID 453 VWFMSFSYGLTV 466

RESULT 14
QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNOEKAIAIATLKRIATEN-----GAP---M 219
ID 094840 PRELIMINARY; PRT; 683 AA.
AC 094840;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0735.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RQ TISSUE=BRAIN;
RQ MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: AB018278; BAA34455.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 683 AA; 77444 MW; B9818120898F09E2 CRC64;

Query Match 23.2%; Score 331; DB 4; Length 683;
Best Local Similarity 27.4%; Pred. No. 1e-15;
Matches 86; Conservative 60; Mismatches 120; Indels 48; Gaps 6;

QY 1 EDAYEATIGFGKFWKLSVLTGLAMMADAMEMMILSILAPQLHCEWRLPSWOVALLTSVVF 60
ID 96 ETIMDECGHGRFQWTLFVLGLALMADGVEFVVGFLVPSAEKDMCLSDSNKMGILVIV 155
QY 61 VGMSSSTLWGNISDOYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG- 119
ID 156 LGMMAGAFILGLADKLGRKRVLSMSLAVNASFASLSSEVQGYGAFGLFCRLISGIGIGA 215
QY 120 VPOSVTLVAEFLPMKARAKCILLIEVFVAIGTVFVWLVAVFVMPSLG----- 166
ID 216 LPIVFAVFEFLSKREKREHLSWLCFMWIGGYASAMAWSIIPHYGWSMGNTNYHPS 275
QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNOEKAIAIATLKRIATEN-----GAPMPLG 222
ID 276 WRVFLVCAFPVFAIGALTTPQESPRFFLENGKHDEAWMLKQVHDTNMRAGHPERV 335
QY 223 KL-----IISROEDRG-----KMRDLTFP-----HFRWTTLLWF 252
ID 336 TVSNIKTPKQMDFTIQTSTGTQWQVGRVRLVFRKTIQKQVNDNALCYVMGPMYRMTLILAV 395
QY 253 IWFSAFYSYGLVL 266
ID 396 VWFMAFYSYGLTV 409
RESULT 15
QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNOEKAIAIATLKRIATEN-----GAP---M 219
ID 094840 PRELIMINARY; PRT; 455 AA.
AC 094840;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PROBABLE MFS TRANSPORTER.
GN PA3467.

Wed Mar 13 13:08:36 2002

```
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stoyer C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.*;
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE004767; AAG06855.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Complete proteome; Transmembrane
SQ SEQUENCE 455 AA; 49357 MW; 2727ED655908A400 CRC64;

Query Match 22.4%; Score 320; DB 2; Length 455;
Best Local Similarity 29.4%; Pred. No. 4.1e-15;
Matches 85; Conservative 58; Mismatches 122; Indels 24; Gaps 6;

Qy 2 DAVEAIGFGKFWKLSVLTGLAWADAMENMILSILAPQLHCEWRLPSHQVALLTSVVFV 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 10 ERLERLPLSPYHRLVFVIALAFEFDSMDLMMTFLLGSIKAEFGLDLSAQAGLIASSFF 69
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 62 GMMSSSTLWGNISDOYGRKTKLIS-VLWTLYYGILSAFAPYVSWILVRLGLVGFIG-G 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 70 GMVIGAAALSGMLADRFGRKPVFOASIVLWGLASVLCSTAGDLDS-LTFYRVLLIGMGME 128
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 120 VPOSVTLYAEFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVMPSLGNRWLLIISAVPLL 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 129 FPIAQSLSEMIPASRRGKYIALMDGEFPLGFAAGCLSYFLLPLTGWRSIFLVLPALPAV 188
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 180 LFAVLCFWLPESARYDVLSGNQEKAIAATLKRIATE-----NGADPMLGKLIISROEDR 232
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 189 FVLAIRFLIPESPRWLEQAGREQADRVLRDIEARVMRSGLGLTELPPPL-----RQPQR 242
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 233 GKMR-----DLFTPHFRWTTLLLWFIWFSNAFSYYGLVLTTELFQ 273
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 243 ERSRPGFFSAFAELWSPAYRRRTLVVWGLWFFALLGYGLTSWLSALLQ 291
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: March 13, 2002, 12:42:30
Job time: 136 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 12:59:50 ; Search time 24.74 Seconds
(without alignments)
1640.750 Million cell up

Title: US-09-911-667A-2
 Perfect score: 548
 Sequence: 1 MEEDLFQLROLPVVKKFRRTG.....NVGRGMHGAGVTRNSGSQE 548

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

```

Word size : 0
Total number of hits satisfying chosen parameters: 192911

```

```
Minimum DB seq length: 100
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```

Database :
1: A_Geneseq_1101.*
2: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
3: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
4: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
5: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
6: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
7: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
8: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
9: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
10: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
11: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
12: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
13: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
14: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
15: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
16: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
17: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
18: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
19: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
20: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
21: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
22: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
23: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|-----------|---|
| 1 | 548 | 100.0 | 548 | 21 | AAV44633 | Human organic cationic amino acid transporter 1 |
| 2 | 7 | 1.3 | 114 | 22 | AAAG90545 | C-glutamylcysteine synthetase |
| 3 | 7 | 1.3 | 135 | 21 | AAG316095 | Zea mays protein f |
| 4 | 7 | 1.3 | 137 | 20 | AAV04945 | Mycobacterium species |
| 5 | 7 | 1.3 | 140 | 20 | AAV41321 | Human secreted protein |
| 6 | 7 | 1.3 | 140 | 20 | AAV04946 | Mycobacterium species |
| 7 | 7 | 1.3 | 153 | 22 | AAG82372 | S. epidermidis open reading frame |
| 8 | 7 | 1.3 | 162 | 22 | AAG73460 | Human gene 4-encod |
| 9 | 7 | 1.3 | 213 | 22 | AA875328 | Gene 44 human secreted protein |
| 10 | 7 | 1.3 | 217 | 22 | AA82731 | S. epidermidis open reading frame |
| 11 | 7 | 1.3 | 230 | 17 | AAH95954 | Eukaryotic cell growth factor |

| | | | | | | |
|----|---|-----|------|----|-----------|-----------------------|
| 12 | 7 | 1.3 | 236 | 21 | AAG30331 | Arabidopsis thaliana |
| 13 | 7 | 1.3 | 257 | 21 | AAW70102 | Staphylococcus aureus |
| 14 | 7 | 1.3 | 309 | 22 | AAW43625 | Human polypeptide |
| 15 | 7 | 1.3 | 329 | 21 | AAG30330 | Arabidopsis thaliana |
| 16 | 7 | 1.3 | 333 | 21 | AAW13689 | C. pneumoniae CR52 |
| 17 | 7 | 1.3 | 333 | 22 | AAG82858 | S. epidermidis ope |
| 18 | 7 | 1.3 | 333 | 22 | AAG83257 | Protein encoded by |
| 19 | 7 | 1.3 | 334 | 20 | AAW35268 | Amino acid sequence |
| 20 | 7 | 1.3 | 346 | 20 | AAV321851 | Human signal pepti |
| 21 | 7 | 1.3 | 346 | 21 | AAG36309 | Arabidopsis thaliana |
| 22 | 7 | 1.3 | 368 | 16 | AAW68961 | Xenorhabdus nemato |
| 23 | 7 | 1.3 | 368 | 20 | AAW97810 | Xenorhabdus nemato |
| 24 | 7 | 1.3 | 396 | 18 | AAW34218 | Streptomyces hydro |
| 25 | 7 | 1.3 | 396 | 19 | AAW55819 | Streptomyces roseo |
| 26 | 7 | 1.3 | 401 | 21 | AAW96284 | Minorhizobium meli |
| 27 | 7 | 1.3 | 420 | 18 | AAW15074 | M. leprae gyrase i |
| 28 | 7 | 1.3 | 430 | 17 | AAW03567 | Arabidopsis thaliana |
| 29 | 7 | 1.3 | 451 | 22 | AAG90061 | C glutamicum prote |
| 30 | 7 | 1.3 | 451 | 22 | AAW76758 | Corynebacterium gl |
| 31 | 7 | 1.3 | 461 | 21 | AAG36308 | Arabidopsis thaliana |
| 32 | 7 | 1.3 | 467 | 20 | AAW85687 | DBX oligosacchari |
| 33 | 7 | 1.3 | 511 | 22 | AAW41958 | Human polypeptide |
| 34 | 7 | 1.3 | 549 | 22 | AAW67471 | Amino acid sequence |
| 35 | 7 | 1.3 | 550 | 22 | AAW06612 | Human protein havi |
| 36 | 7 | 1.3 | 550 | 22 | AAW69091 | Human organic amino |
| 37 | 7 | 1.3 | 599 | 22 | AAW06591 | Human protein havi |
| 38 | 7 | 1.3 | 670 | 19 | AAW29639 | Human secreted pro |
| 39 | 7 | 1.3 | 774 | 19 | AAW80411 | A secreted protein |
| 40 | 7 | 1.3 | 890 | 16 | AAW77845 | Human Rse RPK H |
| 41 | 7 | 1.3 | 890 | 21 | AAW27663 | Human protein PRO8 |
| 42 | 7 | 1.3 | 1266 | 22 | AAW401172 | Human polypeptide |
| 43 | 7 | 1.3 | 1273 | 18 | AAW15078 | M. leprae gyrA pre |
| 44 | 7 | 1.3 | 1297 | 19 | AAW40200 | Infected cell prot |
| 45 | 7 | 1.3 | 1298 | 19 | AAW80810 | The amino acid seq |

ALIGNMENTS

| RESULT | 1 | |
|----------|--|----------------------------|
| AAAY4633 | AAAY4633 standard; Protein; 548 AA. | |
| XX | AAAY44633; | |
| XX | 07-APR-2000 (first entry) | |
| XX | Human organic cation transporter-like protein (OCT1p). | |
| XX | Human; organic cation transporter-like protein; OCT1p; | |
| KW | transmembrane; nootropic; neuroprotective; neuroleptic | |
| KW | antiParkinsonian; antidepressant; cellular process; ce | |
| KW | screen; treatment; prevention; diagnosis; neurodegener | |
| KW | Alzheimer; Parkinson's; Huntington; ALS; amyotrophic l | |
| KW | CNS disorder; central nervous system; schizophrenia; d | |
| KW | behavioural; sleep disorder; Alzheimer's; eating disor | |
| XX | Homo sapiens. | |
| OS | | |
| XX | | |
| Key | Location/Qualifiers | |
| FT | Domain | 1..85 |
| FT | Domain | /label= Cytoplasmic_domain |
| FT | Domain | 141..154 |
| FT | Domain | /label= Cytoplasmic_domain |
| FT | Domain | 199..208 |
| FT | Domain | /label= Cytoplasmic_domain |
| FT | Domain | 259..314 |
| FT | Domain | /label= Cytoplasmic_domain |
| FT | Domain | 395..402 |
| FT | Domain | /label= Cytoplasmic_domain |
| FT | Domain | 448..457 |
| FT | Domain | /label= Cytoplasmic_domain |
| FT | Domain | 511..548 |

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX Corynebacterium glutamicum.
 OS
 XX EP1108790-A2.
 PN 20-JUN-2001.
 PD 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999; 95JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX (KYOWA) KYOWA HAKKO KOGYO KK.
 PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI; 2001-376931/40.
 DR N-PSDB; AAH65764.
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX Claim 17; SEQ ID NO: 4299; 246pp + Sequence Listing; English.
 PS The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX Sequence 114 AA;
 SQ
 Query Match
 Best Local Similarity 100.0%; Score 7; DB 22; Length 114;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 ISVLWTL I61
 Db 58 isvltl 64
 RESULT 3
 ID AAG36095
 AD AAG36095 standard; Protein; 135 AA.
 XX AAG36095;
 AC 18-OCT-2000 (first entry)
 XX Zea mays protein fragment SEQ ID NO: 44187.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence; corn.
 XX Zea mays subsp. mays.
 OS
 XX

PN EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 24-MAY-1999; 99US-0135353.
 PR 25-MAY-1999; 99US-0135629.
 PR 27-MAY-1999; 99US-0136021.
 PR 28-MAY-1999; 99US-0136392.
 PR 01-JUN-1999; 99US-0136782.
 PR 03-JUN-1999; 99US-0137222.
 PR 04-JUN-1999; 99US-0137528.
 PR 07-JUN-1999; 99US-0137502.
 PR 08-JUN-1999; 99US-0137724.
 PR 10-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 14-JUN-1999; 99US-0138847.
 PR 16-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 22-JUN-1999; 99US-0139817.
 PR 23-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 99US-0141842.

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.38; Score 7; DB 21; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 GLVGFGI 187
| | | | |
Db 14 glvgfgl 20

RESULT 4
AAY04945
ID AAY04945 standard; Protein; 137 AA.
XX
AC AAY04945;
XX
DT 06-JUL-1999 (first entry)
XX
DE Mycobacterium species protein sequence 40D.
XX
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW Hybridisation; detection; vaccine; immunisation; infection.
XX
OS Mycobacterium sp.
XX
PN WO9909186-A2.
XX
PD 25-FEB-1999.
XX
PF 14-AUG-1998; 98WO-FR01813.
PR 11-SEP-1997; 97FR-0011325.
PR 14-AUG-1997; 97FR-0010404.

```

XX PA (INSP ) INST PASTEUR.
XX PI Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
XX PI Guigueno A;
XX DR WPI; 1999-181045/15.
XX DR N-PSDB; AAX34198.
XX XX Mycobacterial DNA vectors containing reporter constructs - for
XX PT identifying coding or promoter sequences involved in
XX PT infection-associated protein expression
XX XX Claim 32; Fig 40b; 309pp; French.
XX CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
XX CC proteins from various Mycobacterium species microorganisms. The
XX CC encoding nucleotide sequences can be used as primers and probes for
XX CC methods for detecting and identifying mycobacteria, especially belonging
XX CC to the M. tuberculosis complex. The encoded proteins can be used in
XX CC vaccines for immunisation against a bacterial or viral infection.
XX SQ Sequence 137 AA;

Query Match 1.3%; Score 7; DB 20; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 GLVLLTT 339
Db 116 givlltt 122
|||||

RESULT 5
AAY41321
ID AAY41321 standard; Protein; 140 AA.
XX AC AAY41321;
XX DT 02-DEC-1999 (first entry)
XX DE Human secreted protein encoded by gene 14 clone HMVBS81.
XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX KW developmental abnormality; foetal deficiency; blood; allergy; renal;
XX KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX OS Homo sapiens.
XX XX WO9947540-A1.
XX XX 23-SEP-1999.
XX XX 18-MAR-1999; 99WO-US05804.
XX XX 19-MAR-1998; 98US-0078563.
XX XX 19-MAR-1998; 98US-0078566.
XX XX 19-MAR-1998; 98US-0078573.
XX XX 19-MAR-1998; 98US-0078574.
XX XX 19-MAR-1998; 98US-0078576.
XX XX 19-MAR-1998; 98US-0078577.
XX XX 19-MAR-1998; 98US-0078578.
XX XX 19-MAR-1998; 98US-0078579.
XX XX 01-APR-1998; 98US-0078581.
XX XX 01-APR-1998; 98US-0080312.
XX XX 01-APR-1998; 98US-0080313.
XX XX 01-APR-1998; 98US-0080314.

```

```

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
XX PI Wei Y, Endress CA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
XX PI Olsen HS, Shi Y, Moore PA;
XX XX WPI; 1999-562050/47.
XX DR N-PSDB; AA224824.
XX XX New isolated human genes, useful for diagnosis and treatment of e.g.
XX PT cancers, neurological disorders, immune diseases, inflammation or blood
XX PT disorders -
XX PS Claim 11; Page 366; 484pp; English.
XX CC This sequence represents a secreted human protein encoded by the gene
XX CC clone detailed in the descriptor line. The gene can be used to generate
XX CC fusion proteins by linking to the gene to a human immunoglobulin Fc
XX CC portion (e.g. AA224802) for increasing the stability of the fused
XX CC protein as compared to the human protein only.
XX CC The invention relates to 95 novel genes and their fragments (nucleic
XX CC acid sequences: AA224811-224907; amino acid sequences AAY4108-Y41404)
XX CC which are useful for preventing, treating or ameliorating medical
XX CC conditions e.g. by protein or gene therapy. Also, pathological
XX CC polypeptides can be diagnosed by determining the amount of the new
XX CC polynucleotides in a sample or by determining the presence of mutations in
XX CC the new polynucleotides. Specific uses are described for each of the 95
XX CC polynucleotides, based on which tissues they are most highly expressed in
XX CC (see AA224811 for described uses).
XX SQ Sequence 140 AA;

Query Match 1.3%; Score 7; DB 20; Length 140;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVITGL 92
Db 63 lsvitgl 69
|||||

RESULT 6
AAY04946
ID AAY04946 standard; Protein; 140 AA.
XX AC AAY04946;
XX XX 06-JUL-1999 (first entry)
XX DT 06-JUL-1999 (first entry)
XX DE Mycobacterium species protein sequence 40F.
XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW hybridisation; detection; vaccine; immunisation; infection.
XX XX Mycobacterium sp.
XX XX WO9909186-A2.
XX XX 25-FEB-1999.
XX XX 14-AUG-1998; 98WO-FR01813.
XX XX 11-SEP-1997; 97FR-0011325.
XX XX 14-AUG-1997; 97FR-0010404.
XX XX (INSP ) INST PASTEUR.
XX XX Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
XX PI Guigueno A;
XX DR WPI; 1999-181045/15.

```

DR N-PSDB; AAX34199.
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 XX
 PS Claim 32; Fig 40F; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 XX
 SQ Sequence 140 AA;

Query Match 1.3%; Score 7; DB 20; Length 140;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 GLVLLTT 339
 Db 119 glvlltt 125
 |||||

RESULT 7
 AAG82372
 ID AAG82372 standard; Protein; 153 AA.
 XX
 AC AAG82372;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1838.
 XX
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 DR N-PSDB; AAH53222.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis.
 XX
 PS Claim 18; Page 506; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 XX Sequence 153 AA;

Query Match 1.3%; Score 7; DB 22; Length 153;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 RIATENG 286
 Db 23 riateng 29
 |||||

RESULT 8
 AAG73460
 ID AAG73460 standard; Protein; 162 AA.
 XX
 AC AAG73460;
 XX
 DT 10-AUG-2001 (first entry)
 XX
 DE Human gene 4-encoded secreted protein fragment, SEQ ID NO:235.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; chromosome 11;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angioygenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder; tumour;
 KW endocrine disorder; infection; wound healing; vulnery;
 KW cell culture; chemotaxis; food additive;
 KW binding partner identification.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200134628-A1.
 XX
 PD 17-MAY-2001.
 XX
 PF 08-NOV-2000; 2000WO-US30653.
 XX
 PR 12-NOV-1999; 99US-0164735.
 PR 27-JUL-2000; 2000US-0221193.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Komatsoulis GA, Birse CE, Ni J, Moore PA;
 PI
 XX WPI; 2001-329066/34.
 DR
 XX
 PT Nucleic acids encoding 35 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX
 PS Disclosure; Page 14; 604pp; English.
 XX
 CC AAH32522-AAH32627 represent cDNAs corresponding to 35 human secreted
 CC protein genes, and AAG73346-AAG73448 represent the proteins they encode.
 CC AAG73449-AAG73519 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the

CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC cardiovascular disorders, angioedema, sepsis, diabetes, atherosclerosis,
CC gastrointestinal disorders, pregnancy-related disorders, kidney disorders,
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked,
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.
XX
SQ Sequence 162 AA;

Query Match 1.3%; Score 7; DB 22; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 410 FSPCSLL 416
Db 41 fsfcsll 47

RESULT 9
ID AAB75328 standard; Protein; 213 AA.
AC AAB75328;
DT 03-APR-2001 (first entry)
DE Gene 44 human secreted protein homologous amino acid sequence #147.
KW Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; autoimmune disease; hyperproliferative disorder; cancer;
KW cardiovascular disorder; cerebrovascular disorder; infection;
KW nervous system disorder; ocular disorder; chemotaxis; food additive;
KW secreted protein.
OS Homo sapiens.
XX
XX WO2000077021-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US15135.
XX
XX 11-JUN-1999; 99US-0138632.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI; 2001-071257/08.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

PT Parkinson's diseases and cancers -
PS Disclosure; Page 96; 530pp; English.
XX
CC This invention relates to polynucleotide sequences AAF63789 - AAF63836
CC which encode human secreted proteins AAB75260 - AAB75287. Included in the
CC invention are protein sequences AAB75288 - AAB75341 which are fragments
CC of the secreted proteins and amino acid sequences with which these
CC fragments share homology. Examples of the activities of the proteins and
CC polynucleotides and the activities of their agonists and antagonists
CC include, immunosuppressive; antiarthritic; antirheumatic;
CC antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
CC neurotropic; neuroprotective; antibacterial; virucide; fungicide;
CC ophthalmological; and vulnary activity. The protein and polynucleotide
CC sequences, their agonists and antagonists may be useful for treating,
CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.
XX
SQ Sequence 213 AA;

Query Match 1.3%; Score 7; DB 22; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 LSVLTGL 92
Db 137 lsvltgl 143
RESULT 10
AAG82731
ID AAG82731 standard; Protein; 217 AA.
XX
AC AAG82731;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2556.
XX
XX Staphylococcus epidermidis SK1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
XX
XX N-PSDB; AAH53581.
XX

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 XX useful for vaccinating against infections, e.g. endocarditis -
 PS Claim 18; Page 670; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC *S. epidermidis* polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX

Sequence 217 AA;

Query Match 1.3%; Score 7; DB 22; Length 217;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 280 RIATENG 286
 Db 203 riateng 209
 |||||

RESULT 11

AAH95954
 ID AAR95954 standard; Protein: 230 AA.

AC AAR95954;

DT 24-AUG-1996 (first entry)

DE Eukaryotic cell growth inhibiting factor.

KW Eukaryotic cell growth inhibiting factor; ageing; cancer; infection;
 KW gene therapy; *Schizosaccharomyces pombe*; fission yeast.

OS Homo sapiens.

PN WO9617933-A2.

PD 13-JUN-1996.

PF 05-DEC-1995; 95WO-JP02488.

PR 02-JUN-1995; 95JP-0136252.

PR 09-DEC-1994; 94JP-0306602.

PR 16-MAR-1995; 95JP-0057716.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Igarashi K, Sasada R, Takeyama M;

DR WPI; 1996-287178/29.

DR N-PSDB; AAT30564.

PT DNA encoding a eukaryotic cell growth inhibiting factor, and its

prod. - useful as anticancer and infection treatment

XX

PS Claim 20; Page 50-51; 73pp; English.

XX A eukaryotic cell growth inhibiting factor (AAR95954) was identified
 CC as the product of cDNA clone pTB1848 (AAT30564). This clone was
 CC isolated by incorporating cDNA from aged normal human diploid fibroblast
 CC MKC-5 cells into vector pTB1589 under control of the nmt1 promoter,
 CC transforming of *Schizosaccharomyces pombe* cells and screening for
 CC transformants showing cDNA-dependent growth inhibition. Other cDNA
 CC clones coding for eukaryotic cell growth inhibiting factors were
 CC also obtd. (see also AAR95951-57). These proteins are useful as
 CC anticancer agents and infection remedies. They inhibit nucleic
 CC acid synthesis in target cells. They can also be used to terminate
 CC the cell cycle of cultured cells at a given time point.
 XX

Sequence 230 AA;

Query Match 1.3%; Score 7; DB 17; Length 230;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 71 EDAVEAI 77
 Db 122 edaveai 128
 |||||

RESULT 12

AAG30331

ID AAG30331 standard; Protein: 236 AA.

XX AAG30331;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 36241.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.3%; Score 7; DB 21; Length 236;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 EDAAVEAI 77
Db 27 edaveai 33

RESULT 13
AAV70102
ID AAV70102 standard; Protein: 257 AA.

XX AC AAV70102;
XX DT 05-JUN-2000 (first entry)
XX DE Staphylococcal enterotoxin A.

XX KW Superantigen toxin; SAg; Staphylococcal enterotoxin A; SEA; cytostatic;
KW antibacterial; vaccine; MHC class II receptor; T-cell antigen receptor;
KW antibody; toxoid; staphylococcal/streptococcal toxin; diagnosis;
KW treatment; superantigen-associated bacterial infection.

XX OS Staphylococcus sp.

XX FH Key Location/Qualifiers
FT Peptide 2..25
FT /label= Leader_peptide
FT Protein 26..253
FT /label= Mature_Staphylococcal_enterotoxin_A
FT /note= "Includes transcription start site residue, Met"
FT Misc-difference 42
FT /note= "Encoded by TTG"
FT Misc-difference 125
FT /note= "Encoded by CCA"

XX WO200009154-A1.
XX PD 24-FEB-2000.

XX PF 13-AUG-1998; 98WO-US16766.
XX PR 13-AUG-1998; 98WO-US16766.
XX PA (REED-) REED ARMY INST RES WALTER.

XX PI Ulrich RG, Olson MA, Bavari S;
XX DR WPI: 2000-224177/19.
XX DR N-PSDB; AAZ51105.

XX PT Nucleic acid encoding superantigen toxin useful as a vaccine and for
PT diagnosis of superantigen-associated bacterial infections -

XX PS Claim 7; Page 72-73; 118pp; English.

XX CC The present amino acid sequence is the Staphylococcal enterotoxin A
CC (SEA), a bacterial superantigen toxin (SAg). The coding region
CC of the SAg toxin when altered by site directed mutagenesis, results
CC in disruption of binding of the toxin to both the MHC class II or T-cell
CC antigen receptor. SEA has antibacterial and cytostatic activity. This
CC sequence is useful for the production of SEA vaccines and specific

CC antibodies. This vaccine overcomes the disadvantages of the chemically
CC inactivated toxoids and is designed to protect individuals against one
CC or several related staphylococcal and streptococcal toxins. It is used
CC for the diagnosis and treatment or amelioration of superantigen.
CC associated bacterial infections.

XX SQ Sequence 257 AA;

Query Match 1.3%; Score 7; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 427 TLLLFIA 433
Db 7 tlllfia 13

RESULT 14
AAM43625
ID AAM43625 standard; Protein: 309 AA.

XX AC AAM43625;

XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 303.

XX KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nontropic; neurop; dective; antibacterial; virucide;
KW fungicide; ophthalmological; cytos; tic; immunosuppressive; nontropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.

XX OS Homo sapiens.

XX PN WO200155308-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01309.

XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488781/53.
DR N-PSDB; AA163931.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
XX Claim 11; SEQ ID NO 303; 664pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AA1634497-AA163660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 309 AA;

Query Match 1.3%; Score 7; DB 22; Length 309;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | | | |
|----|-----|-------------|----|--------------|---------------|
| QY | 211 | LLIEVFw 217 | PR | 18-JUN-1999; | 99US-0139454. |
| | | | PR | 18-JUN-1999; | 99US-0139455. |
| Db | 264 | llievfw 270 | PR | 18-JUN-1999; | 99US-0139456. |
| | | | PR | 18-JUN-1999; | 99US-0139457. |
| | | | PR | 18-JUN-1999; | 99US-0139458. |
| | | | PR | 18-JUN-1999; | 99US-0139459. |
| | | | PR | 18-JUN-1999; | 99US-0139460. |
| | | | PR | 18-JUN-1999; | 99US-0139461. |
| | | | PR | 18-JUN-1999; | 99US-0139462. |
| | | | PR | 18-JUN-1999; | 99US-0139463. |
| | | | PR | 18-JUN-1999; | 99US-0139750. |
| | | | PR | 18-JUN-1999; | 99US-0139763. |
| | | | PR | 21-JUN-1999; | 99US-0139817. |
| | | | PR | 22-JUN-1999; | 99US-0139899. |
| | | | PR | 23-JUN-1999; | 99US-0140353. |
| | | | PR | 23-JUN-1999; | 99US-0140354. |
| | | | PR | 24-JUN-1999; | 99US-0140695. |
| | | | PR | 28-JUN-1999; | 99US-0140823. |
| | | | PR | 29-JUN-1999; | 99US-0140991. |
| | | | PR | 30-JUN-1999; | 99US-0141287. |
| | | | PR | 01-JUL-1999; | 99US-0141842. |
| | | | PR | 01-JUL-1999; | 99US-0142154. |
| | | | PR | 02-JUL-1999; | 99US-0142055. |
| | | | PR | 06-JUL-1999; | 99US-0142390. |
| | | | PR | 08-JUL-1999; | 99US-0142803. |
| | | | PR | 09-JUL-1999; | 99US-0142920. |
| | | | PR | 12-JUL-1999; | 99US-0142977. |
| | | | PR | 13-JUL-1999; | 99US-0143542. |
| | | | PR | 14-JUL-1999; | 99US-0143624. |
| | | | PR | 15-JUL-1999; | 99US-0144005. |
| | | | PR | 16-JUL-1999; | 99US-0144085. |
| | | | PR | 16-JUL-1999; | 99US-0144086. |
| | | | PR | 19-JUL-1999; | 99US-0144325. |
| | | | PR | 19-JUL-1999; | 99US-0144331. |
| | | | PR | 19-JUL-1999; | 99US-0144332. |
| | | | PR | 19-JUL-1999; | 99US-0144333. |
| | | | PR | 19-JUL-1999; | 99US-0144334. |
| | | | PR | 19-JUL-1999; | 99US-0144335. |
| | | | PR | 20-JUL-1999; | 99US-0144352. |
| | | | PR | 20-JUL-1999; | 99US-0144632. |
| | | | PR | 20-JUL-1999; | 99US-0144884. |
| | | | PR | 21-JUL-1999; | 99US-0144814. |
| | | | PR | 21-JUL-1999; | 99US-0145086. |
| | | | PR | 21-JUL-1999; | 99US-0145088. |
| | | | PR | 22-JUL-1999; | 99US-0145085. |
| | | | PR | 22-JUL-1999; | 99US-0145087. |
| | | | PR | 22-JUL-1999; | 99US-0145089. |
| | | | PR | 22-JUL-1999; | 99US-0145192. |
| | | | PR | 23-JUL-1999; | 99US-0145145. |
| | | | PR | 23-JUL-1999; | 99US-0145218. |
| | | | PR | 23-JUL-1999; | 99US-0145224. |
| | | | PR | 26-JUL-1999; | 99US-0145276. |
| | | | PR | 27-JUL-1999; | 99US-0145913. |
| | | | PR | 27-JUL-1999; | 99US-0145918. |
| | | | PR | 27-JUL-1999; | 99US-0145919. |
| | | | PR | 28-JUL-1999; | 99US-0145951. |
| | | | PR | 02-AUG-1999; | 99US-0146386. |
| | | | PR | 02-AUG-1999; | 99US-0146388. |
| | | | PR | 02-AUG-1999; | 99US-0146389. |
| | | | PR | 03-AUG-1999; | 99US-0147038. |
| | | | PR | 04-AUG-1999; | 99US-0147204. |
| | | | PR | 04-AUG-1999; | 99US-0147302. |
| | | | PR | 05-AUG-1999; | 99US-0147192. |
| | | | PR | 05-AUG-1999; | 99US-0147260. |
| | | | PR | 06-AUG-1999; | 99US-0147303. |
| | | | PR | 06-AUG-1999; | 99US-0147416. |
| | | | PR | 09-AUG-1999; | 99US-0147493. |
| | | | PR | 09-AUG-1999; | 99US-0147935. |
| | | | PR | 10-AUG-1999; | 99US-0148171. |
| | | | PR | 11-AUG-1999; | 99US-0148319. |
| | | | PR | 12-AUG-1999; | 99US-0148341. |
| | | | PR | 13-AUG-1999; | 99US-0148565. |
| | | | PR | 13-AUG-1999; | 99US-0148684. |

RESULT 15
AAG30330
ID AAG30330 standard; Protein; 329 AA.
AC AAG30330;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 36240.
XX
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

Job time: 86 sec

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.3%; Score 7; DB 21; Length 329;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 71 EDAVEAI 77
| | | | |
Db 120 edaveai 126

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 12:59:50 ; Search time 14.07 Seconds
(without alignments)
876.461 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 548
Sequence: 1 MEEDLFOLRLPVVKFRRTG.....MVGRGMIGAGVTRNSGSQE 548

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 49859

Minimum DB seq length: 100
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep:*

2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep:*

3: /cgn2.6/ptodata/2/1aa/6A_COMB.pep:*

4: /cgn2.6/ptodata/2/1aa/6B_COMB.pep:*

5: /cgn2.6/ptodata/2/1aa/PctUS_COMB.pep:*

6: /cgn2.6/ptodata/2/1aa/backfiles1.pep:*

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 7 | 1.3 | 115 | 3 | US-08-513-974B-351 |
| 2 | 7 | 1.3 | 257 | 3 | US-08-486-099-113 |
| 3 | 7 | 1.3 | 257 | 3 | US-08-360-107A-123 |
| 4 | 7 | 1.3 | 257 | 3 | US-08-484-223B-113 |
| 5 | 7 | 1.3 | 257 | 3 | US-08-919-597-113 |
| 6 | 7 | 1.3 | 257 | 3 | US-08-475-668A-113 |
| 7 | 7 | 1.3 | 257 | 3 | US-08-485-551A-113 |
| 8 | 7 | 1.3 | 257 | 3 | US-08-471-913A-113 |
| 9 | 7 | 1.3 | 257 | 4 | US-08-485-264A-113 |
| 10 | 7 | 1.3 | 278 | 2 | US-08-569-168-2 |
| 11 | 7 | 1.3 | 383 | 2 | US-08-569-168-7 |
| 12 | 7 | 1.3 | 430 | 1 | US-08-601-435-2 |
| 13 | 7 | 1.3 | 430 | 2 | US-08-931-047-2 |
| 14 | 7 | 1.3 | 430 | 2 | US-08-783-202-2 |
| 15 | 7 | 1.3 | 890 | 1 | US-08-445-640-2 |
| 16 | 7 | 1.3 | 890 | 3 | US-08-170-558-2 |
| 17 | 7 | 1.3 | 890 | 3 | US-08-447-314-2 |
| 18 | 7 | 1.3 | 890 | 3 | US-08-445-461-2 |
| 19 | 7 | 1.3 | 911 | 1 | US-08-286-305A-1 |
| 20 | 7 | 1.3 | 911 | 2 | US-08-441-104A-1 |
| 21 | 7 | 1.3 | 911 | 2 | US-08-440-816A-1 |
| 22 | 7 | 1.3 | 911 | 4 | US-09-417-381A-1 |
| 23 | 7 | 1.3 | 1298 | 2 | US-08-690-473-2 |
| 24 | 7 | 1.3 | 1298 | 4 | US-09-259-821A-2 |
| 25 | 7 | 1.3 | 1298 | 4 | US-08-843-659-2 |
| 26 | 7 | 1.3 | 3457 | 2 | US-08-416-603-4 |
| 27 | 6 | 1.1 | 102 | 2 | US-08-710-749-21 |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

US-08-513-974B-351

; Sequence 351, Application US/08513974B

; Patent No. 6114139

; GENERAL INFORMATION:

; APPLICANT: Hinuma, Shuji

; APPLICANT: Hosoya, Masaki

; APPLICANT: Fujii, Ryo

; APPLICANT: Ohtaki, Tetsuya

; APPLICANT: Fukusumi, Shoji

; APPLICANT: Ogi, Kazuhiro

; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN.

; NUMBER OF SEQUENCES: 380

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

; STREET: 130 Water Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/513,974B

; FILING DATE: 14-SEP-1995

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP95/01599

; FILING DATE: 10-AUG-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 7-093989

; FILING DATE: 19-AUG-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 7-057186

; FILING DATE: 16-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 7-007177

; FILING DATE: 20-JAN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 6-326611

; FILING DATE: 28-DEC-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 6-270017

; FILING DATE: 02-NOV-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 6-236357

; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-236356
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-189274
; FILING DATE: 11-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-189273
; FILING DATE: 11-AUG-1945
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-189272
; FILING DATE: 11-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S.
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 45753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 351:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-513-974B-351

Query Match 1.3% Score 7: DB 3: Length 115;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 90 TGLAWMA 96
Db 15 TGLAWMA 21

RESULT 2
US-08-486-099-113
; Sequence 113, Application US/08486099
; Patent No. 6013263
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
; TITLE OF INVENTION: B VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.

; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-486-099-113

Query Match 1.3% Score 7: DB 3: Length 257;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 427 TLLLFIA 433
Db 7 TLLLFIA 13

RESULT 3
US-08-360-107A-123
; Sequence 123, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 123:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-360-107A-123

Query Match 1.3%; Score 7; DB 3; Length 257;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433
 Db 7 TLLLFIA 13

RESULT 4

US-08-484-223B-113
 ; Sequence 113, Application US/08484223B
 ; Patent No. 6020459
 ; GENERAL INFORMATION:
 ; APPLICANT: Bolognesi, Dani P.
 ; APPLICANT: Matthews, Thomas J.
 ; APPLICANT: Wild, Carl T.
 ; APPLICANT: Barney, Shawn O.
 ; APPLICANT: Lambert, Dennis M.
 ; APPLICANT: Petteway, Stephen R.
 ; APPLICANT: Langlois, Alphonse J.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
 ; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
 ; NUMBER OF SEQUENCES: 245
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 07-JUN-1995
 ; APPLICATION NUMBER: US/08/484,223B
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coruzzi, Laura A.
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7872-029
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 113:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 257 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-484-223B-113

Query Match 1.3%; Score 7; DB 3; Length 257;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433
 Db 7 TLLLFIA 13

RESULT 5

US-08-919-597-113
 ; Sequence 113, Application US/08919597

Patent No. 6054265
 ; GENERAL INFORMATION:
 ; APPLICANT: Bolognesi, Dani P.
 ; APPLICANT: Matthews, Thomas J.
 ; APPLICANT: Wild, Carl T.
 ; APPLICANT: Barney, Shawn O.
 ; APPLICANT: Lambert, Dennis M.
 ; APPLICANT: Petteway, Stephen R.
 ; APPLICANT: Langlois, Alphonse J.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
 ; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
 ; NUMBER OF SEQUENCES: 273
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: US/08/919,597
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/470,896
 ; FILING DATE: 06-JUN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coruzzi, Laura A.
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7872-020
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 113:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 257 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-919-597-113

Query Match 1.3%; Score 7; DB 3; Length 257;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433
 Db 7 TLLLFIA 13

RESULT 6

US-08-475-668A-113
 ; Sequence 113, Application US/08475668A
 ; Patent No. 6060065
 ; GENERAL INFORMATION:
 ; APPLICANT: Barney, Shawn O.
 ; APPLICANT: Lambert, Dennis M.
 ; APPLICANT: Petteway, Stephen R.
 ; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
 ; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
 ; NUMBER OF SEQUENCES: 211
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT: Bolognesi, Dani P.
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-026
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-475-668A-113

Query Match 1.3%; Score 7; DB 3; Length 257;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433
|||||
DB 7 TLLLFIA 13

RESULT 7
US-08-485-551A-113
; Sequence 113, Application US/08485551A
; Patent No. 6068973
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
; FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICANT: Bolognesi, Dani P.
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-023
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-551A-113

Query Match 1.3%; Score 7; DB 3; Length 257;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433
|||||
DB 7 TLLLFIA 13

RESULT 8
US-08-471-913A-113
; Sequence 113, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICANT: Bolognesi, Dani P.
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-030
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown

MOLECULE TYPE: protein
US-08-471-913A-113

Query Match 1.3%; Score 7; DB 3; Length 257;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 TLLLFIA 433
Db 7 TLLLFIA 13

RESULT 9

US-08-485-264A-113
Sequence 113, Application US/08485264A
Patent No. 6228983
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,264A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-021
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-264A-113

Query Match 1.3%; Score 7; DB 4; Length 257;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 TLLLFIA 433
Db 7 TLLLFIA 13

RESULT 10

US-08-569-168-2
Sequence 2, Application US/08569168
Patent No. 5972687
GENERAL INFORMATION:
APPLICANT: Smigielski, Adam J.
APPLICANT: Akhurst, Raymond J.
TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lowe, Price, Leblanc & Becker
STREET: 99 Canal Center Plaza, Suite 300
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,168
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Price, Robert L.
REGISTRATION NUMBER: 22,685
REFERENCE/DOCKET NUMBER: 1451-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-684-1111
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-168-2

Query Match 1.3%; Score 7; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 240 LLLLSAV 246
Db 190 LLLLSAV 196

RESULT 11

US-08-569-168-7
Sequence 7, Application US/08569168
Patent No. 5972687
GENERAL INFORMATION:
APPLICANT: Smigielski, Adam J.
APPLICANT: Akhurst, Raymond J.
TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lowe, Price, Leblanc & Becker
STREET: 99 Canal Center Plaza, Suite 300
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,168
FILING DATE:

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Price, Robert L.
; REGISTRATION NUMBER: 22,685
; REFERENCE/DOCKET NUMBER: 1451-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-168-7

```

```

Query Match      1.3%; Score 7; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 240 LLILSAV 246
   | | | | |
Db 190 LLILSAV 196

```

RESULT 12

```

; Sequence 2, Application US/08601435-2
; Patent No. 5759801
; GENERAL INFORMATION:
; APPLICANT:

```

```

; TITLE OF INVENTION: DNA sequence coding for a protein
; of A, thaliana having a delta-5,7 sterol,delta-7
; TITLE OF INVENTION: reductase activity, delta7-Red protein, production
; of A, thaliana having a delta-5,7 sterol,delta-7
; TITLE OF INVENTION: reductase activity, delta7-Red protein, production
; of A, thaliana having a delta-5,7 sterol,delta-7
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/601,435
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: FR 9501723
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9506517
; FILING DATE: 01-JUN-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-601-435-2

```

```

Query Match      1.3%; Score 7; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 425 VLTLLLF 431
   | | | | |
Db 390 VLTLLLF 396

```

RESULT 13

```

; US-08-931-047-2

```

```

; Sequence 2, Application US/08931047
; Patent No. 5965417
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA sequence coding for a protein of
; A, thaliana having a delta-5,7 sterol,
; TITLE OF INVENTION: delta-7 reductase activity, delta7-Red
; protein, production process, strains
; TITLE OF INVENTION: of transformed yeasts, uses.
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,047
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9501723
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION NUMBER: FR 9506517
; FILING DATE: 01-JUN-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-931-047-2

```

```

Query Match      1.3%; Score 7; DB 2; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 425 VLTLLLF 431
   | | | | |
Db 390 VLTLLLF 396

```

RESULT 14

```

; US-08-783-202-2
; Sequence 2, Application US/08783202
; Patent No. 5989881
; GENERAL INFORMATION:
; APPLICANT:

```

```

; TITLE OF INVENTION: DNA sequence coding for a protein of A.
; thaliana having a delta-5,7 sterol,delta-7 reductase
; TITLE OF INVENTION: activity, delta7-Red protein, production
; process, strains
; TITLE OF INVENTION: of transformed yeasts, uses.
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/783,202
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9501723
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION NUMBER: FR 9506517
; FILING DATE: 01-JUN-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid

```

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-783-202-2

Query Match 1.3%; Score 7; DB 2; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 425 VLTLLLF 431
Db 390 VLTLLLF 396

RESULT 15
US-08-445-640-2
: Sequence 2, Application US/08445640
: Patent No. 5709858
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J.
: APPLICANT: Mark, Melanie R.
: APPLICANT: Scadden, David T.
: APPLICANT: Baker, Kevin P.
: APPLICANT: Baron, Will F.
: TITLE OF INVENTION: Protein Tyrosine Kinases
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/445,640
: FILING DATE: 22-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/170558
: FILING DATE: 20-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/157563
: FILING DATE: 23-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 854C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 890 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-08-445-640-2

Query Match 1.3%; Score 7; DB 1; Length 890;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 499 LLAALAS 505
Db 26 LLAALAS 32

Search completed: March 7, 2002, 13:01:37
Job time: 107 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 12:59:50 ; Search time 19.42 Seconds
(without alignments)
2149.519 Million cell updates/sec

Title: US-09-911-667a-2
Perfect score: 548
Sequence: 1 MEEDLQLRLPVVKPRRTG.....MVGGMHGAGVTRNSGSGE 548

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 189327

Minimum DB seq length: 100
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : JIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 110 | 20.1 | 144 | 2 | T50622 | hypothetical prote |
| 2 | 19 | 3.5 | 300 | 2 | S15786 | glucose transport |
| 3 | 19 | 3.5 | 529 | 2 | T23190 | hypothetical prote |
| 4 | 9 | 1.6 | 331 | 2 | S2163 | sucrose specific r |
| 5 | 9 | 1.6 | 331 | 2 | E5878 | sucrose specific t |
| 6 | 8 | 1.5 | 236 | 2 | D69048 | conserved hypothet |
| 7 | 8 | 1.5 | 293 | 2 | B83783 | hypothetical prote |
| 8 | 8 | 1.5 | 359 | 2 | A42462 | acetoin catabolism |
| 9 | 8 | 1.5 | 413 | 1 | S34305 | nodulation protein |
| 10 | 8 | 1.5 | 413 | 2 | JC5178 | probable starvatio |
| 11 | 8 | 1.5 | 425 | 2 | F70608 | hypothetical prote |
| 12 | 8 | 1.5 | 472 | 2 | T34748 | transmembrane tran |
| 13 | 8 | 1.5 | 826 | 2 | F85510 | probable acyl-CoA |
| 14 | 8 | 1.5 | 826 | 2 | F64746 | probable membrane |
| 15 | 8 | 1.5 | 1256 | 2 | S60461 | gene flightless-I |
| 16 | 7 | 1.3 | 105 | 2 | T18029 | hypothetical prote |
| 17 | 7 | 1.3 | 137 | 2 | H70547 | hypothetical prote |
| 18 | 7 | 1.3 | 141 | 1 | HATJD | hemoglobin alpha-D |
| 19 | 7 | 1.3 | 141 | 2 | G70144 | hypothetical prote |
| 20 | 7 | 1.3 | 145 | 1 | PSKF2U | phospholipase A2 (|
| 21 | 7 | 1.3 | 178 | 2 | D71451 | phospholipase A2 (|
| 22 | 7 | 1.3 | 178 | 2 | G71195 | hypothetical prote |
| 23 | 7 | 1.3 | 180 | 2 | G82858 | hypothetical prote |
| 24 | 7 | 1.3 | 190 | 2 | T09136 | ADP-ribosylation f |
| 25 | 7 | 1.3 | 192 | 2 | S43563 | R01H10.5 protein - |
| 26 | 7 | 1.3 | 194 | 1 | HSTR1 | histone H1 - trout |
| 27 | 7 | 1.3 | 194 | 2 | D72242 | conserved hypothet |
| 28 | 7 | 1.3 | 200 | 2 | S67524 | calcium-binding pr |
| 29 | 7 | 1.3 | 206 | 1 | HSTRIR | histone H1 - rainb |

ALIGNMENTS

RESULT 1

T50622
hypothetical protein DKF2p761H039.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50622
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A:Reference number: 225145
A:Accession: T50622
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-144 <AAA>
A:Cross-references: EMBL:AL359592
A:Experimental source: adult amygdala; clone DKF2p761H039
C:Genetics:
A:Note: DKF2p761H039.1

Query Match 20.1% Score 110; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 5.1e-103;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LCFVIFSCSLLLFCVGRNVLTLLFTARAFISGFOAAVYVTPVPTATRALGLGTC 464
|||||
Db 1 LCFVIFSCSLLLFCVGRNVLTLLFTARAFISGFOAAVYVTPVPTATRALGLGTC 60
QY 465 SCMARVGALITPPIAQVMLESSVYLTAVYSCCLLAAALASCFLPETK 514
|||||
Db 61 SCMARVGALITPPIAQVMLESSVYLTAVYSCCLLAAALASCFLPETK 110

RESULT 2

S15786
glucose transport protein homolog - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C:Date: 06-Jan-1995 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: S15786
R:Craxton, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; H
Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.
submitted to the EMBL Data Library, May 1991
A:Reference number: S15786
A:Accession: S15786
A:Molecule type: DNA
A:Residues: 1-300 <CRA>
A:Cross-references: EMBL:Z11115
C:Genetics:
A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3
C:Superfamily: Caenorhabditis elegans glucose transport protein
Query Match 3.5% Score 19; DB 2; Length 300;

conserved hypothet
nonstructural prot
nonstructural prot
nonstructural prot
hypothetical prote
conserved hypothet
H+-transporting AT
hypothetical prote
photosynthetic rea
enterotoxin A prec
probable proteasom
flagellar biosynth
molybdate-binding
probable beta-lact
hypothetical prote
conserved hypothet

Best Local Similarity 100.0%; Pred. No. 7.5e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GFGIGGVQSVTLTLYAEFLP 202
|||||
Db 183 GFGIGGVQSVTLTLYAEFLP 201

RESULT 3
T23190
hypothetical protein ZK637.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
C:Accession: T23190
R:Craxton, M.
submitted to the EMBL Data Library, April 1993
A:Reference number: Z19704
A:Accession: T23190
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-529 <WII>
A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1
C:Genetics:
A:Gene: CESP:ZK637.1
A:Map position: 3
A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 3.5%; Score 19; DB 2; Length 529;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GFGIGGVQSVTLTLYAEFLP 202
|||||
Db 183 GFGIGGVQSVTLTLYAEFLP 201

RESULT 4
S52163
sucrose specific repressor - Escherichia coli
C:Species: Escherichia coli
C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 08-Oct-1999
C:Accession: S52163
R:Bockmann, J.
submitted to the EMBL Data Library, September 1994
A:Reference number: S52160
A:Accession: S52163
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <HOC>
A:Cross-references: EMBL:X81461; NID:g608705; PIDN:CAA57220.1; PID:g608709
C:Superfamily: lac repressor

Query Match 1.6%; Score 9; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LCFWLPESA 262
|||||
Db 181 LCFWLPESA 189

RESULT 5
E85878
sucrose specific transcription regulator [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85878
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85878
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <STO>
A:Cross-references: GB:AE005174; NID:gi2516731; PIDN:AA057489.1; GSPDB:GN00145; UWGF
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3626

Query Match 1.6%; Score 9; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LCFWLPESA 262
|||||
Db 181 LCFWLPESA 189

RESULT 6
D69048
conserved hypothetical protein MTH1364 - Methanobacterium thermoautotrophicum (strain
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: D69048
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H:
A:Reference number: A69000; MUID:98037514
A:Accession: D69048
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-236 <MTH>
A:Cross-references: GB:AE000899; GB:AE000666; NID:g2622468; PIDN:AAB85841.1; PID:g26
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1364

Query Match 1.5%; Score 8; DB 2; Length 236;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 RGLVGFGI 187
|||||
Db 82 RGLVGFGI 89

RESULT 7
B83783
hypothetical protein BH1066 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: B83783
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
A:Reference number: A83650; MUID:20263314
A:Accession: B83783
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04785.1; GSPDB:
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1066

Query Match 1.5%; Score 8; DB 2; Length 293;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 LLLFAVLC 255
 |||||
 Db 20 LLLFAVLC 27

RESULT 8
 A42462
 acetoin catabolism protein AcoX - Alcaligenes eutrophus (strain H16)
 C:Species: Alcaligenes eutrophus
 C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 30-Sep-1993
 C:Accession: A42462
 R: Priefert, H.; Hein, S.; Krueger, N.; Zeh, K.; Schmidt, B.; Steinbuechel, A.
 J. Bacteriol. 173, 4056-4071, 1991
 A:Title: Identification and molecular characterization of the Alcaligenes eutrophus H16
 A:Reference number: A42462; MUID:91286190
 A:Accession: A42462
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <PR1>
 A:Cross-references: GB:M66060

Query Match 1.5%; Score 8; DB 2; Length 359;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 LLAALASC 506
 |||||
 Db 49 LLAALASC 56

RESULT 9
 S34305
 nodulation protein nodC - Rhizobium sp.
 C:Species: Rhizobium sp.
 C:Date: 13-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 16-Jul-1999
 C:Accession: S34305
 R: Relix, B.; Perret, X.; Golinowsky, W.; Pueppke, S.G.; Krishnan, H.B.; Broughton, W.J.
 submitted to the EMBL Data Library, June 1993
 A:Description: Lipo-oligosaccharide Nod-factor signals permit rhizobial penetration into
 A:Reference number: S34303
 A:Accession: S34305
 A:Molecule type: DNA
 A:Residues: 1-413 <REL>
 A:Cross-references: EMBL:X73362; NID:g312347; PIDN:CAA51774.1; PID:g312350
 C:Comment: This is one of the proteins, coded by nodulation genes, that are required for
 C:Genetics:
 A:Gene: nodC
 C:Superfamily: nodulation protein nodC
 C:Keywords: nodulation

Query Match 1.5%; Score 8; DB 1; Length 413;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVLTGLA 93
 |||||
 Db 319 LSVLTGLA 326

RESULT 10
 JC5178
 probable starvation-sensing protein A - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 16-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 31-Jan-2000
 C:Accession: JC5178; T42059
 R: Schneider, D.; Bruton, C.J.; Chater, K.F.
 Gene 177, 243-251, 1996

Query Match 1.5%; Score 8; DB 2; Length 413;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402
 |||||
 Db 70 DRLGRKKT 77

RESULT 12
 T34748
 transmembrane transport protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
 C:Accession: T34748
 R: Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, August 1998
 A:Reference number: Z21556
 A:Accession: T34748
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-472 <MUR>
 A:Cross-references: EMBL:AL031184; PIDN:CAA20171.1; GSPDB:GN00070; SCOEDB:SC2A11.02c
 A:Experimental source: strain A3(2)
 C:Genetics:

A:Title: Characterization of spaA, a Streptomyces coelicolor gene homologous to a gen
 A:Reference number: JC5178; MUID:97080529
 A:Accession: JC5178
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-413 <SCH>
 A:Cross-references: EMBL:X94190; NID:g1694907; PIDN:CAA63900.1; PID:e222100; PID:g169
 C:Comment: This protein is involved in an intercellular signalling system.

Query Match 1.5%; Score 8; DB 2; Length 413;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 VELDDGAA 55
 |||||
 Db 383 VELDDGAA 390

RESULT 11
 F70608
 hypothetical protein Rv1200 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: F70608
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70608
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-425 <COL>
 A:Cross-references: GB:Z93777; GB:AL123456; NID:g3261726; PIDN:CAB07823.1; PID:g19290
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv1200
 C:Superfamily: citrate utilization determinant

Query Match 1.5%; Score 8; DB 2; Length 425;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402
 |||||
 Db 70 DRLGRKKT 77

RESULT 12
 T34748
 transmembrane transport protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
 C:Accession: T34748
 R: Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, August 1998
 A:Reference number: Z21556
 A:Accession: T34748
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-472 <MUR>
 A:Cross-references: EMBL:AL031184; PIDN:CAA20171.1; GSPDB:GN00070; SCOEDB:SC2A11.02c
 A:Experimental source: strain A3(2)
 C:Genetics:

A:Gene: SCOEDB:SC2Alli.02c

C:Superfamily: citrate utilization determinant

Query Match 1.5%; Score 8; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 395 DRLGRKKT 402

Db 86 DRLGRKKT 93

RESULT 13

F85510
probable acyl-CoA dehydrogenase (EC 1.3.99.-) yafH [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: F85510
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85510

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-826 <STO>

A:Cross-references: GB:AE0051174; NID:g12512989; PIDN:AAG54546.1; GSPDB:GN00145; UWGP:Z02
A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yafH

C:Keywords: oxidoreductase

Query Match 1.5%; Score 8; DB 2; Length 826;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 MMILSILA 108

Db 13 MMILSILA 20

RESULT 14

F64746

probable membrane protein b0221 - Escherichia coli

N:Alternate names: hypothetical protein b0221

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999

C:Accession: F64746

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64746

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-826 <BLAT>

A:Cross-references: GB:AE000130; GB:U00096; NID:g1786402; PIDN:AAC73325.1; PID:g1786414;

A:Experimental source: strain K-12, substrain MG1655

F:10-26/Domain: transmembrane #status predicted <TM1>

F:33-49/Domain: transmembrane #status predicted <TM2>

F:51-67/Domain: transmembrane #status predicted <TM3>

F:292-308/Domain: transmembrane #status predicted <TM4>

Query Match

1.5%; Score 8; DB 2; Length 826;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 MMILSILA 108

Db 13 MMILSILA 20

RESULT 15

S60461

gene flightless-I protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000

C:Accession: S60461; T08425

R:de Couet, H.G.; Fong, K.S.K.; Weeds, A.G.; McLaughlin, P.J.; Miklos, G.L.G.

Genetics 141, 1049-1059, 1995

A:Title: Molecular and mutational analysis of a gelsolin-family member encoded by the

A:Reference number: S60461; MUID:96129280

A:Accession: S60461

A:Molecule type: DNA

A:Residues: 1-1256 <DEC>

A:Cross-references: GB:AF017777; EMBL:U28044; NID:g3004652; PIDN:AAC28407.1; PID:g3030

R:Maleszka, R.; de Couet, H.G.; Miklos, G.L.

Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998

A:Title: Data transferability from model organisms to human beings: insights from the

A:Reference number: Z16415; MUID:98188272

A:Accession: T08425

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1256 <NAL>

A:Cross-references: GB:AF017777; GB:U80043; GB:U28044; NID:g3004652; PIDN:AAC28407.1;

C:Genetics:

A:Gene: fli-I; flightless-I

A:Cross-references: FlyBase:FBgn000070

A:Introns: 18/3; 1070/2; 1120/3

C:Superfamily: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:53-75/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:76-100/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:101-123/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

F:124-147/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

F:148-170/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>

F:171-193/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

F:220-242/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>

F:243-265/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>

F:266-288/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>

F:497-830/Domain: gelsolin repeat homology <GEL1>

F:892-1250/Domain: gelsolin repeat homology <GEL2>

Query Match

1.5%; Score 8; DB 2; Length 1256;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LFQLRQLP 12

Db 184 LFQLRQLP 191

Search completed: March 7, 2002, 13:02:04

Job time: 134 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 13:00:50 ; Search time 13.29 seconds
(without alignments)
1511.838 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 548
Sequence: 1 MEEDIQLRLPLVVKFRRTG.....MVGGMHGAGVTRNSGSGUE 548

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 88480

Minimum DB seq length: 100

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 19 | 3.5 | 529 | 1 Y0U1_CAEEL | P10638 caenorhabdi |
| 2 | 9 | 1.6 | 331 | 1 CSCR_ECOLI | P40715 escherichia |
| 3 | 8 | 1.5 | 359 | 1 AC0X_ALCEU | P27748 alceguen |
| 4 | 8 | 1.5 | 413 | 1 N0DC_RHISN | P50357 rhizobium s |
| 5 | 8 | 1.5 | 1256 | 1 FLII_DROME | P24020 drosophila |
| 6 | 7 | 1.3 | 141 | 1 HBAD_SPHPU | P10062 spheonodon p |
| 7 | 7 | 1.3 | 145 | 1 PA20_BUNMU | P00606 bungarus mu |
| 8 | 7 | 1.3 | 194 | 1 HI_SALTR | P02234 salmo trutt |
| 9 | 7 | 1.3 | 202 | 1 CC23_ORCLI | P80363 oncorhynch |
| 10 | 7 | 1.3 | 206 | 1 H1_ONCMY | P06350 oncorhynch |
| 11 | 7 | 1.3 | 229 | 1 VP8_BTIV10 | P08363 bluetongue |
| 12 | 7 | 1.3 | 229 | 1 VP8_BTIV1 | Q04684 bluetongue |
| 13 | 7 | 1.3 | 229 | 1 VP8_BTIV13 | Q04685 bluetongue |
| 14 | 7 | 1.3 | 229 | 1 VP8_BTIV17 | Q04686 bluetongue |
| 15 | 7 | 1.3 | 229 | 1 VP8_BTIV1A | P13841 bluetongue |
| 16 | 7 | 1.3 | 229 | 1 VP8_BTIV1S | P23705 bluetongue |
| 17 | 7 | 1.3 | 229 | 1 VP8_BTIV2A | Q04687 bluetongue |
| 18 | 7 | 1.3 | 242 | 1 ATP1_ODOSI | Q00825 odontella s |
| 19 | 7 | 1.3 | 255 | 1 RCXM_RHOPH | P51751 rhodospiril |
| 20 | 7 | 1.3 | 257 | 1 ETXA_STAAU | P13163 staphylococ |
| 21 | 7 | 1.3 | 258 | 1 PSMA_AERPE | Q9yc01 aeropyrum p |
| 22 | 7 | 1.3 | 292 | 1 NLA_DROME | Q9x218 drosophila |
| 23 | 7 | 1.3 | 306 | 1 YDFC_BACSU | P36680 bacillus su |
| 24 | 7 | 1.3 | 315 | 1 YK04_CAEEL | P34272 caenorhabdi |
| 25 | 7 | 1.3 | 319 | 1 POBB_PSEPS | Q52186 pseudomonas |
| 26 | 7 | 1.3 | 323 | 1 RCXM_RHOVI | P06010 rhodopseudo |
| 27 | 7 | 1.3 | 324 | 1 HEM2_VIBCH | Q9ktb6 vibrio chol |
| 28 | 7 | 1.3 | 347 | 1 C5AR_MOUSE | P30993 mus musculu |
| 29 | 7 | 1.3 | 356 | 1 PCBI_HUMAN | Q15365 homo sapien |
| 30 | 7 | 1.3 | 356 | 1 PCBI_RABIT | O19048 oryctolagus |
| 31 | 7 | 1.3 | 373 | 1 FLHF_AQUAE | O67266 aquifex aeo |
| 32 | 7 | 1.3 | 381 | 1 GSPF_KLEPN | P15745 klebsiella |
| 33 | 7 | 1.3 | 415 | 1 SPHS_SYNP7 | P39664 synechococc |

RESULT 1

| ID | Y0U1_CAEEL | STANDARD: | PRT: | 529 AA. |
|----|--|-----------|------|---------|
| AC | P30638: Q21101; | | | |
| DT | 01-APR-1993 (Rel. 25, Created) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | |
| DE | HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III. | | | |
| GN | ZK637.1. | | | |
| OS | Caenorhabditis elegans. | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; | | | |
| OC | Rhabditidae; Peloderinae; Caenorhabditis. | | | |
| OX | NCBI_TaxID=6239; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=BRISTOL N2; | | | |
| RC | MEDLINE=92168156; PubMed=1538779; | | | |
| RA | Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R., | | | |
| RA | Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A., | | | |
| RA | Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T., | | | |
| RA | Ainscough R., Waterston R.; | | | |
| RT | *The C. elegans genome sequencing project: a beginning.*; | | | |
| RL | Nature 356:37-41(1992). | | | |
| RN | [2] | | | |
| RP | REVIEWS. | | | |
| RC | STRAIN=BRISTOL N2; | | | |
| RA | Durbin R.; | | | |
| RL | Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE). | | | |
| CC | -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; Z11115; CAA77460.1; - | | | |
| DR | EMBL; Z22175; CAA77460.1; JOINED. | | | |
| DR | EMBL; Z22175; CAA80131.1; - | | | |
| DR | EMBL; Z11115; CAA80131.1; JOINED. | | | |
| DR | PIR; S15786; S15786 | | | |
| DR | WormPep; ZK637.1; CE066638. | | | |
| DR | InterPro; IPR003662; sub_transportr. | | | |
| DR | Pfam; PF00083; sugar_tr; 1. | | | |
| DR | PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG. | | | |
| DR | PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG. | | | |
| KW | Hypothetical protein; Transmembrane; Transport. | | | |
| FT | Transmem 86 106 POTENTIAL. | | | |
| FT | Transmem 122 142 POTENTIAL. | | | |
| FT | Transmem 158 178 POTENTIAL. | | | |
| FT | Transmem 238 258 POTENTIAL. | | | |
| FT | Transmem 320 340 POTENTIAL. | | | |
| FT | Transmem 373 393 POTENTIAL. | | | |

P24151 rhizobium l
Q51955 pseudomonas
P28364 eupioten oc
P15663 influenza a
P48019 oryza sativ
P47533 mycoplasma
Q9pjp6 chlamydia m
P21186 mumps virus
P42359 streptococc
P45758 escherichia
P53118 saccharomyc
P41639 pinus thunb

ALIGNMENTS

FT TRANSMEM 411 431 POTENTIAL.
 FT TRANSMEM 482 502 POTENTIAL.
 SQ SEQUENCE 529 AA; 58317 MW; 8D2FF4CBA15ECD2D CRC64;

Query Match 3.5%; Score 19; DB 1; Length 529;
 Best Local Similarity 100.0%; Pred. No. 1.9e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GFGIGGVPOSVTLYAEFLP 202
 DB 183 GFGIGGVPOSVTLYAEFLP 201

RESULT 2
 ID CSCR_ECOLI STANDARD; PRT; 331 AA.
 AC P40715;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SUCROSE OPERON REPRESSOR (CSC OPERON REGULATORY PROTEIN).
 GN CSCR.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EC3132;
 RA Bockmann J.;

RL Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: REPRESSOR FOR THE CSC OPERON. BINDS D-FRUCTOSE AS
 CC AN INDUCER (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; X81461; CAA57220.1; -.
 DR HSP; P03023; LCSD.
 DR InterPro; IPR000843; HTH_LacI.
 DR InterPro; IPR001761; Peripla_BP_like.
 DR Pfam; PF00356; lacI; 1.
 DR Pfam; PF00532; Peripla_BP_like; 1.
 DR PRINTS; PRO0036; HTHLACI.
 DR SMART; SM00354; HTH_LACI; 1.
 DR PROSITE; PS00356; HTH_LACI_FAMILY; 1.
 KW Transcription regulation; DNA-binding; Repressor.
 FT DNA_BIND 4 23 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 331 AA; 36470 MW; BF9AEA07C10B431F CRC64;

Query Match 1.6%; Score 9; DB 1; Length 331;
 Best Local Similarity 100.0%; Pred. No. 0.75;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LCFWLPESA 262
 DB 181 LCFWLPESA 189

RESULT 3
 ID ACOX_ALCEU STANDARD; PRT; 359 AA.
 AC P27748;
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE ACETOIN CATABOLISM PROTEIN X.
 GN ACOX.
 OS Alcaligenes eutrophus (Ralstonia eutropha).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HI6;
 RX MEDLINE=91286190; PubMed=2061286;
 RA Priefert H., Hein S., Krueger N., Zeh K., Schmidt B.,
 RA Steinbuechel A.;
 RT "Identification and molecular characterization of the Alcaligenes
 RT eutrophus HI6 aco operon genes involved in acetoin catabolism.";
 RL J. Bacteriol. 173:4056-4071(1991).
 CC -!- FUNCTION: ESSENTIAL FOR ACETOIN CATABOLISM.
 CC -!- PATHWAY: ACETOIN CATABOLISM.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE (POTENTIAL).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; M66060; AAA21947.1; -.
 KW Membrane.
 SQ SEQUENCE 359 AA; 37934 MW; 6825322F5B8C50E5 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 359;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 LLAALASC 506
 DB 49 LLAALASC 56

RESULT 4
 ID NODC_RHISN STANDARD; PRT; 413 AA.
 AC P50357;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.-) (MODULATION PROTEIN C).
 GN NODC OR Y4HG.
 OS Rhizobium sp. (strain NGR234).
 OG Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed=9163424;
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL Nature 387:394-401(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95075295; PubMed=7984092;
 RA Relic B., Perret X., Estrada-Garcia M.T., Kopicinska J., Golinowski W.,
 RA Krishnan H.B., Pueppke S.G., Broughton W.J.;
 RT "Nod factors of Rhizobium are a key to the legume door.";
 RL Mol. Microbiol. 13:171-178(1994).
 CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF NOD FACTOR, A SULFATED
 CC N-ACYL-BETA-1,4-TETRASACCHARIDE OF N-ACETYLGLUCOSAMINE WHICH

CC AND D. HB A IS A TETRAMER OF TWO ALPHA-A AND TWO BETA-1, HB A' IS
CC A TETRAMER OF TWO ALPHA-A AND TWO BETA-2, HB D IS A TETRAMER OF
CC TWO ALPHA-D AND TWO BETA-2.
CC MISCELLANEOUS: SPHENODON'S HBS HAVE PROPERTIES NOT FOUND IN OTHER
CC REPTILES: POOR COOPERATIVITY, HIGH AFFINITY FOR OXYGEN, SMALL BOHR
CC AND HALDANE EFFECTS, APPRECIABLE PHOSPHATE EFFECTS (THOSE
CC PROPERTIES ARE ALSO FOUND IN THE HBS OF PRIMITIVE URODELE AND
CC CAECILIAN AMPHIBIANS).
CC -1- SIMILARITY: HIGH DEGREE OF SIMILARITY WITH ALPHA D SEQUENCE FROM
CC BIRDS AND TURTLE.
CC PIR: S01137; HATJD.
CC HSP: P01958; 2MHB.
CC InterPro: IPR002338; Alpha_haem.
CC InterPro: IPR000971; Globin.
CC Pfam: PF00042; globin; 1.
CC PRINTS: PR00612; ALPHAHAEM.
CC PROSITE: PS01033; GLOBIN; 1.
CC Heme: Oxygen transport; Respiratory protein; Erythrocyte.
CC FT METAL 58 IRON (HEME DISTAL LIGAND).
CC FT METAL 87 IRON (HEME PROXIMAL LIGAND).
CC SQ SEQUENCE 141 AA; 16272 MW; F5B8E633C9F9AA1 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 FEVVLAV 229
Db 105 FEVVLAV 111
|||||

RESULT 7
ID PA20_BUNMU STANDARD; PRT: 145 AA.
AC P00606;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PHOSPHOLIPASE A2 PRECURSOR (EC 3.1.1.4) (PHOSPHATIDYLCHOLINE 2-
DE ACYLHYDROLASE).
OS Bungarus multicinctus (Many-banded krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Bungarinae; Bungarus.
OX NCBI_TaxID=8616;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=90356415; PubMed=2388841;
RA Danse J.M.;
RT "Nucleotide sequence encoding for non-toxic phospholipase-A2 from
RT Bungarus multicinctus.";
RL Nucleic Acids Res. 18:4608-4608(1990).
RN [2]
RP SEQUENCE OF 28-145.
RC TISSUE=Venom;
RX MEDLINE=81168081; PubMed=7217037;
RA Kondo K., Toda H., Narita K.;
RT "Amino acid sequence of phospholipase A from Bungarus multicinctus
RT venom.";
RL J. Biochem. 89:37-47(1981).
CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE
CC 2-ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERYL-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

DR EMBL; U01182; AAC03566.1; -;
DR EMBL; AF017777; AAC28407.1; -;
DR EMBL; AF132184; AAD34772.1; -;
DR EMBL; AF003568; AAF50830.1; ALT_SEQ.
DR HSP; P02640; 2VIL.
DR FlyBase; Fgn0000709; f111.
DR InterPro: IPR001974; Gelsolin.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_TYP.
DR Pfam; PF00626; Gelsolin; 4.
DR Pfam; PF00560; LRR; 12.
DR PRINTS; PR00019; LEURICHRPT.
DR PRINTS; PR00597; GELSOLIN.
DR SMART; SM00262; GEL; 6.
DR SMART; SM00370; LRR; 3.
DR SMART; SM00369; LRR_TYP; 2.
KW Developmental protein; Repeat; Leucine-rich repeat; Polymorphism.
FT REPEAT 4 28
FT REPEAT 29 51
FT REPEAT 52 74
FT REPEAT 75 99
FT REPEAT 100 122
FT REPEAT 124 145
FT REPEAT 147 169
FT REPEAT 171 192
FT REPEAT 218 241
FT REPEAT 243 264
FT REPEAT 265 287
FT REPEAT 289 312
FT REPEAT 313 335
FT REPEAT 336 358
FT REPEAT 360 381
FT REPEAT 499 557
FT REPEAT 746 789
FT REPEAT 1064 1102
FT REPEAT 1165 1206
FT VARIANT 601 601
FT CONFLICT 1067 1070 STFC -> HYFS (IN REF. 5).
FT CONFLICT 1068 1068 T -> A (IN REF. 2).
SQ SEQUENCE 1256 AA; 143681 MW; CF0056EFAA88DB92 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1256;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LFQLRQLP 12
Db 184 LFQLRQLP 191
|||||

RESULT 6
ID HBAD_SPHPU STANDARD; PRT: 141 AA.
AC P10062;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HEMOGLOBIN ALPHA-D CHAIN.
OS Sphenodon punctatus (Hatteria) (Tuatara).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Rhynchocephalia; Sphenodontidae; Sphenodon.
OX NCBI_TaxID=8508;
RN [1]
RP SEQUENCE.
RX MEDLINE=89105321; PubMed=3214555;
RA Abbasi A., Wells R.M.G., Brittain T., Braunitzer G.;
RT "Primary structure of the hemoglobins from Sphenodon (Sphenodon
RT punctatus, Tuatara, Rhynchocephalia). Evidence for the expression of
RT alpha D-gene.";
RL Biol. Chem. Hoppe-Seyler 369:755-764(1988).
CC -1- SUBUNIT: THERE ARE THREE FORMS OF HEMOGLOBIN IN SPHENODON: A, A'

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53406; CAA37482.1; -.
DR PIR: S10981; PSKF20.
DR HSSP: P15445; I43F.
DR InterPro: IPR001211; P1P_A2.
DR Pfam: PF00068; Phoslip; 1.
DR PRINTS: PR00389; PHPLIPASEA2.
DR ProDom: PD000303; PLP_A2; 1.
DR SMART: SM00085; PA2c; 1.
DR PROSITE: PS00118; PA2_HIS; 1.
DR PROSITE: PS00119; PA2_ASP; 1.
DR Hydrolase; Lipid degradation; Calcium; Multigene family; Signal;
KW Venom.
FT SIGNAL 1 27
FT CHAIN 28 145 PHOSPHOLIPASE A2.
FT ACT_SITE 73 73 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT ACT_SITE 38 97 BY SIMILARITY.
FT DISULFID 52 144 BY SIMILARITY.
FT DISULFID 54 70 BY SIMILARITY.
FT DISULFID 69 125 BY SIMILARITY.
FT DISULFID 76 118 BY SIMILARITY.
FT DISULFID 86 111 BY SIMILARITY.
FT DISULFID 104 116 BY SIMILARITY.
FT CA_BIND 74 74 BY SIMILARITY.
FT CA_BIND 145 145 AA; 15593 MW; F7959376589967CA CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 240 LLILSAV 246
DB 6 LLILSAV 12
RESULT 8
HL_SALTR STANDARD; PRT; 194 AA.
AC P02254;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H1.
OS Salmo trutta (Brown trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8032;
RN [1]
RP SEQUENCE
RX MEDLINE=78023898; PubMed=913397;
RA McLeod A.R., Wong N.C.W., Dixon G.H.;
RT "The amino-acid sequence of trout-testis histone H1.";
RL Eur. J. Biochem. 78:281-291(1977).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR: A02583; HSTR1.
DR HSSP: P08287; ICHC.
DR InterPro: IPR001386; Linker_histone.
DR Pfam: PF00538; linker_histone; 1.
DR SMART: SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation; Phosphorylation.
FT MOD_RES 1 1 ACETYLATION (IN 90% OF THE CHAINS).
FT MOD_RES 145 145 PHOSPHORYLATION.
FT MOD_RES 161 161 PHOSPHORYLATION.
FT MOD_RES 182 182 PHOSPHORYLATION.
FT
us-09-911-667a-2.oligo.rsp
Page 5
```

```

-!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CELLS.
-!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL: A28981; AAA42840.1; -
PIR: A29153; P8XR10.
PIR: S10543; S10543.
InterPro: IPR002565; Orbi_NS3.
Pfam: PF01616; Orbi_NS3; 1.
ProDom: PD003183; Orbi_NS3; 1.
Nonstructural protein.
CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
SEQUENCE 229 AA: 25602 MW: 1A23B53198ECB684 CRC64;
-----
Query Match 1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 122 VALLTSV 128
| | | | |
Db 130 VALLTSV 136
-----
RESULT 12
VP8_BTv11 ID VP8_BTv11 STANDARD; PRT; 229 AA.
AC Q04684;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
DE S10.
GN Bluetongue virus (serotype 11 / isolate USA).
OS Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OC NCBI_TaxID=33716;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295710; PubMed=1318624;
RA Hwang G.-Y., Yang Y.-Y., Chiu J.-F., Li J.K.-K.;
RT "Sequence conservation among the cognate nonstructural NS3/3A protein
RT genes of six bluetongue viruses.";
RL Virus Res. 23:151-161(1992).
-!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CELLS.
-!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL: L08631; AAA42835.1; -
InterPro: IPR002565; Orbi_NS3.
Pfam: PF01616; Orbi_NS3; 1.
ProDom: PD003183; Orbi_NS3; 1.
Nonstructural protein.
CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
SEQUENCE 229 AA: 25374 MW: 22695FC2331DED61 CRC64;
-----
QY 122 VALLTSV 128
| | | | |
Db 130 VALLTSV 136
-----

```

```

Query Match          1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 VALLTSV 128
    |||||
Db 130 VALLTSV 136

RESULT 13
VP8_BTIV13
ID VP8_BTIV13 STANDARD; PRT; 229 AA.
AC Q04685;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 13 / isolate USA).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=33717;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295710; PubMed=1318624;
RA Huang G.-Y., Yang Y.-Y., Chou J.-F., Li J.K.-K.;
RT "Sequence conservation among the cognate nonstructural NS3/3A protein
RT genes of six bluetongue viruses.";
RL Virus Res. 23:151-161(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08629; AAA42836.1; -
CC InterPro: IPR002565; Orbi_NS3.
CC Pfam: PF01616; Orbi_NS3; 1.
CC ProDom: PD003183; Orbi_NS3; 1.
CC Nonstructural protein.
CC CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
CC CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
CC SEQUENCE 229 AA; 25507 MW; 3FDFDIA3138335B8 CRC64;

Query Match          1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 VALLTSV 128
    |||||
Db 130 VALLTSV 136

RESULT 14
VP8_BTIV17
ID VP8_BTIV17 STANDARD; PRT; 229 AA.
AC Q04686;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 17 / isolate USA).

```

```

OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=33718;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295710; PubMed=1318624;
RA Huang G.-Y., Yang Y.-Y., Chou J.-F., Li J.K.-K.;
RT "Sequence conservation among the cognate nonstructural NS3/3A protein
RT genes of six bluetongue viruses.";
RL Virus Res. 23:151-161(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08630; AAA42837.1; -
CC InterPro: IPR002565; Orbi_NS3.
CC Pfam: PF01616; Orbi_NS3; 1.
CC ProDom: PD003183; Orbi_NS3; 1.
CC Nonstructural protein.
CC CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
CC CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
CC SEQUENCE 229 AA; 25499 MW; 3D99853D39A56CBA CRC64;

Query Match          1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 VALLTSV 128
    |||||
Db 130 VALLTSV 136

RESULT 15
VP8_BTIV1A
ID VP8_BTIV1A STANDARD; PRT; 229 AA.
AC P13841; Q65733;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 1 / isolate Australia).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=10904;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88187687; PubMed=2833571;
RA Gould A.R.;
RT "Nucleotide sequence of the Australian bluetongue virus serotype 1
RT RNA segment 10.";
RL J. Gen. Virol. 69:945-949(1988).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00253; BAA00184.1; -

```

DR EMBL; D00253; BAA00185.1; -.
DR PIR; A28600; P8XRAU.
DR InterPro; IPR002565; Orbi_NS3.
DR Pfam; PF01616; Orbi_NS3; 1.
DR ProDom; PD003183; Orbi_NS3; 1.
KW Nonstructural protein.
FT CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
FT CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
SQ SEQUENCE 229 AA; 25501 MW; 3132BDCE86C4325 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 VALLTSV 128
|
|
|
|
|
|
|
Db 130 VALLTSV 136

Search completed: March 7, 2002, 13:02:58
Job time: 128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 13:00:30 ; Search time 29.03 seconds
(without alignments)
2761.186 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 548
Sequence: 1 MEEDFOLRLPVVKFRRTG.....MVGKMHGAGVTRNSGSGQ: 548

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0
Total number of hits satisfying chosen parameters: 382221

Minimum DB seq length: 100
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-----------|--------------------|
| 1 | 110 | 20.1 | 144 | 4 Q9NPW5 | Q9npw5 homo sapien |
| 2 | 98 | 17.9 | 548 | 11 Q9Z217 | Q9z217 rattus norv |
| 3 | 13 | 2.4 | 497 | 5 Q9W1D4 | Q9w1d4 drosophila |
| 4 | 8 | 1.5 | 205 | 8 Q78791 | Q78791 osmerus mor |
| 5 | 8 | 1.5 | 236 | 1 Q27417 | Q27417 methanobact |
| 6 | 8 | 1.5 | 293 | 2 Q9KD25 | Q9kd25 bacillus ha |
| 7 | 8 | 1.5 | 413 | 2 P95726 | P95726 streptomyc |
| 8 | 8 | 1.5 | 425 | 2 O05301 | O05301 mycobacteri |
| 9 | 8 | 1.5 | 452 | 2 Q53254 | Q53254 rhizobium t |
| 10 | 8 | 1.5 | 472 | 2 O86563 | O86563 streptomyce |
| 11 | 8 | 1.5 | 814 | 2 P71283 | P71283 escherichia |
| 12 | 8 | 1.5 | 1019 | 10 P94078 | P94078 arabidopsis |
| 13 | 8 | 1.5 | 1060 | 5 Q917Q3 | Q917q3 drosophila |
| 14 | 7 | 1.3 | 101 | 11 Q9CZK2 | Q9czk2 mus musculu |
| 15 | 7 | 1.3 | 103 | 10 Q9ZRN3 | Q9zrn3 galega orie |
| 16 | 7 | 1.3 | 103 | 12 Q9QNV8 | Q9qnv8 human immun |
| 17 | 7 | 1.3 | 105 | 12 Q98577 | Q98577 paramecium |
| 18 | 7 | 1.3 | 107 | 12 Q9WL65 | Q9wl65 human immun |
| 19 | 7 | 1.3 | 121 | 12 Q9Q7U5 | Q9q7u5 human immun |

| | | | | | |
|----|---|-----|-----|-----------|--------------------|
| 20 | 7 | 1.3 | 132 | 9 Q9XJP0 | Q9xjp0 bacterioph |
| 21 | 7 | 1.3 | 132 | 10 Q9FMJ8 | Q9fmj8 arabidopsis |
| 22 | 7 | 1.3 | 133 | 5 Q9GX97 | Q9gx97 leishmania |
| 23 | 7 | 1.3 | 137 | 2 O06415 | O06415 mycobacteri |
| 24 | 7 | 1.3 | 141 | 2 O51335 | O51335 borrelia bu |
| 25 | 7 | 1.3 | 152 | 11 Q9D2C5 | Q9d2c5 mus musculu |
| 26 | 7 | 1.3 | 154 | 11 Q9CWB2 | Q9cwb2 caenorhabdi |
| 27 | 7 | 1.3 | 157 | 5 Q9N3B8 | Q9n3b8 caenorhabdi |
| 28 | 7 | 1.3 | 161 | 2 Q9CPK8 | Q9cpk8 pasteurella |
| 29 | 7 | 1.3 | 169 | 3 Q9HFP3 | Q9hfp3 cladosporiu |
| 30 | 7 | 1.3 | 175 | 2 P95548 | P95548 pseudomonas |
| 31 | 7 | 1.3 | 178 | 1 O58005 | O58005 pyrococcus |
| 32 | 7 | 1.3 | 178 | 1 O59518 | O59518 pyrococcus |
| 33 | 7 | 1.3 | 178 | 5 Q9GVQ2 | Q9gvq2 leishmania |
| 34 | 7 | 1.3 | 179 | 2 O07077 | O07077 bordetella |
| 35 | 7 | 1.3 | 180 | 2 Q9A3M1 | Q9a3m1 caulobacter |
| 36 | 7 | 1.3 | 183 | 2 Q9A3M1 | Q9a3m1 caulobacter |
| 37 | 7 | 1.3 | 190 | 5 Q76227 | Q76227 trypanosoma |
| 38 | 7 | 1.3 | 194 | 2 Q9X1M2 | Q9x1m2 thermotoga |
| 39 | 7 | 1.3 | 194 | 2 Q915B5 | Q915b5 streptococc |
| 40 | 7 | 1.3 | 199 | 2 Q9X545 | Q9x545 corynebacte |
| 41 | 7 | 1.3 | 209 | 5 Q21621 | Q21621 caenorhabdi |
| 42 | 7 | 1.3 | 216 | 12 Q9WLP6 | Q9wlp6 bluetongue |
| 43 | 7 | 1.3 | 216 | 12 Q9WLP5 | Q9wlp5 bluetongue |
| 44 | 7 | 1.3 | 216 | 12 Q9WLP4 | Q9wlp4 bluetongue |
| 45 | 7 | 1.3 | 216 | 12 Q98WJ9 | Q98wj9 bluetongue |

ALIGNMENTS

RESULT 1
Q9NPW5 ID Q9NPW5 PRELIMINARY; PRT; 144 AA.
AC Q9NPW5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 15.4 KDA PROTEIN (FRAGMENT).
GN DKEZP761H039.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=AMYGDALA;
RC Blum H., Bauersachs S., Mewes H.W., Weil B., Wiemann S.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL359592; CAB94878.1; -
DR Hypothetical protein.
KW NON_TER
FT SEQUENCE 144 AA; 15387 MW; 769A51A9564FFFCDC CRC64;

Query Match 20.1%; Score 110; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.4e-101;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 405 LCFVIFSCSLLLFTCVGRNVLTLLFIARAFISGGFOAAVYTPVPTATRALGLGTC 464
DB 1 LCFVIFSCSLLLFTCVGRNVLTLLFIARAFISGGFOAAVYTPVPTATRALGLGTC 60

QY 465 SGMARVGALITPFIQVWLESSVYLTAVYSGCCLLAALASCFPIETKG 514
DB 61 SGMARVGALITPFIQVWLESSVYLTAVYSGCCLLAALASCFPIETKG 110

RESULT 2
Q9Z217 ID Q9Z217 PRELIMINARY; PRT; 548 AA.
AC Q9Z217;
DT 01-MAY-1994 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SV2 RELATED PROTEIN.
 GN SVOP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RX TISSUE=Brain.
 RC MEDLINE=99019745; PubMed=9801366;
 RX Janz R., Hofmann K., Sudhof T.C.;
 RA "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests
 RT novel transport functions of synaptic vesicles.";
 RL J. Neurosci. 18:9269-9281(1998).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 DR EMBL: AF060173; AAC78627.1;
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Transmembrane.
 SQ SEQUENCE 548 AA; 60804 MW; 19AD8475B7579496 CRC64;

 Query Match 17.9%; Score 98; DB 11; Length 548;
 Best Local Similarity 100.0%; Pred. No. 7.7e-89;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 251 FAVLCFWLPESARYDVLGNQEKATIKRIATENGAMPGLKLIISROEDRGKMRDLPT 310
 DB 251 FAVLCFWLPESARYDVLGNQEKATIKRIATENGAMPGLKLIISROEDRGKMRDLPT 310
 QY 311 PFRWTTLLWFWNSAFSYGLVLLTTELFQAGDVC 348
 DB 311 PFRWTTLLWFWNSAFSYGLVLLTTELFQAGDVC 348

 RESULT 3
 Q9W1D4 PRELIMINARY; PRT; 497 AA.
 ID Q9W1D4
 AC Q9W1D4;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CG4324 PROTEIN.
 GN CG4324.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RX TISSUE=Brain.
 RC MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champ E., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spaden-Klam A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 CC EMBL: AE003462; AAF47135.1;
 DR FlyBase: FBgn0034956; CG4324.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Transmembrane.
 SQ SEQUENCE 497 AA; 55360 MW; 678F44A3286B7BD4 CRC64;

 Query Match 2.4%; Score 13; DB 5; Length 497;
 Best Local Similarity 100.0%; Pred. No. 0.00043;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 190 VPQSVTLYAERLP 202
 DB 181 VPQSVTLYAERLP 193

 RESULT 4
 Q78791 PRELIMINARY; PRT; 205 AA.
 ID Q78791
 AC Q78791;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (FRAGMENT).
 OS Osmerus mordax (Rainbow smelt).
 OG MitoChondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Osmeriformes; Osmeridae; Osmerus.
 ON NCBI_TaxID=8014;
 RX TISSUE=Brain.
 RC MEDLINE=20196006; PubMed=10731132;
 RX Pigeon D., Dodson J.J., Bernatchez L.;
 RA "A mtDNA analysis of spatio-temporal distribution of two genetically
 RT distinct sympatric larval populations of rainbow smelt (Osmerus
 RT mordax) in the middle estuary of the St. Lawrence River, Quebec,
 RT Canada.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
 CC (BY SIMILARITY).
 CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: TO THE ATPASE A CHAIN FAMILY.
 CC EMBL: AF034748; AAC24026.1;
 DR InterPro: IPR000568; ATP_synt_A.

DR Pfam: PF00119; ATP-synt_A; 1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASE_A; 1.
 KW C(=O); Hydrogen ion transport; Mitochondrion; Transmembrane.
 FT NON_TER 1
 SQ SEQUENCE 205 AA; 22598 MW; AE66B1DAB77A984C CRC64;

Query Match 1.5%; Score 8; DB 8; Length 205;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 122 VALLTSVV 129
 Db 167 VALLTSVV 174
 |||||

RESULT 5
 O27417 PRELIMINARY; PRT; 236 AA.
 AC O27417;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CONSERVED PROTEIN.
 GN MTH1364.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanothermobacter.
 OX NCBI_TaxID=145262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., DuBois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L.C., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicari R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics."
 RL J. Bacteriol. 179:7135-7155(1997).
 DR EMBL: AE000899; AAB85841.1; -
 DR InterPro: IPR003748; DUF169.
 DR Pfam: PF02596; DUF169; 1.
 KW Complete proteome.
 SQ SEQUENCE 236 AA; 26177 MW; 81B09F4EB67E6BFA CRC64;

Query Match 1.5%; Score 8; DB 1; Length 236;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 180 RGLVGFGI 187
 Db 82 RGLVGFGI 89
 |||||

RESULT 6
 Q9KDD5 PRELIMINARY; PRT; 293 AA.
 AC Q9KDD5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BH1066 PROTEIN.
 GN BH1066.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001510; BAB04785.1; -
 DR InterPro: IPR000515; BPD_transp.
 DR Pfam: PF00528; BPD_transp; 1.
 KW Complete proteome.
 SQ SEQUENCE 293 AA; 33010 MW; A2B8A16E6217BD56 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 293;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 LLLFAVLC 255
 Db 20 LLLFAVLC 27
 |||||

RESULT 7
 P95726 PRELIMINARY; PRT; 413 AA.
 AC P95726;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE SPAA (PROBABLE STARVATION SENSING PROTEIN).
 GN SPAA.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97080529; PubMed=8921874;
 RA Schneider D., Bruton C.J., Chater K.F.;
 RT "Characterization of spaA, a Streptomyces coelicolor gene homologous
 RT to a gene involved in sensing starvation in Escherichia coli."
 RL Gene 1177:243-251(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: X94190; CAA63900.1; -
 DR EMBL: AL450350; CAC16962.1; -
 DR InterPro: IPR001354; MR_MLE.
 DR Pfam: PF01188; MR_MLE; 1.
 DR PROSITE: PS00908; MR_MLE_1; 1.
 SQ SEQUENCE 413 AA; 45206 MW; 7EDC6183BC8DE36F CRC64;

```

Query Match      1.5%; Score 8; DB 2; Length 413;
Best Local Similarity 100.0%; Pred. No. 35;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 VELDDGAA 55
DB 383 VELDDGAA 390
|||||||

RESULT 8
Q05301 PRELIMINARY; PRT; 425 AA.
ID O05301;
AC O05301;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 44.6 KDA PROTEIN.
GN RV1200 OR MTCI364.12
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.F., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala K.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DE EMBL: Z93777; CAB07823.1; -.
DR TubercuList; RV1200; -.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Complete proteome; Hypothetical protein; Transmembrane.
SQ SEQUENCE 425 AA; 44575 MW; 67D44D03F0EA7256 CRC64;

Query Match      1.5%; Score 8; DB 2; Length 425;
Best Local Similarity 100.0%; Pred. No. 36;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402
DB 70 DRLGRKKT 77
|||||||

RESULT 9
Q53254 PRELIMINARY; PRT; 452 AA.
ID Q53254;
AC Q53254;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NODA, NODB, NOD C, AND NODD1 GENES.
GN NODC.
OS Rhizobium tropici.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CFN299;
RX MEDLINE=97084572; PubMed=8930915;
RA Debell F., Plazanel C., Roche P., Pujol C., Savagnac A., Rosenberg C.,
RA Prome J.C., Denariet J.;
RT "The Noda proteins of Rhizobium meliloti and Rhizobium tropici specify
RT the N-acylation of Nod factors by different fatty acids.";
RL Mol. Microbiol. 22:303-314(1996).
DR EMBL: X98514; CAA67139.1; -.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1. GAMMA; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
SQ SEQUENCE 452 AA; 50019 MW; 2FA3683DCE27A603 CRC64;

Query Match      1.5%; Score 8; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 38;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVLTGLA 93
DB 319 LSVLTGLA 326
|||||||

RESULT 10
O86563 PRELIMINARY; PRT; 472 AA.
ID O86563;
AC O86563;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRANSMEMBRANE TRANSPORT PROTEIN.
GN SC2A11.02C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC Parkhill J., Barrell B.G., Rajandream M.A.;
RA Murphy L., Harris D.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: AL031184; CAA20171.1; -.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR003880; Phosphatant_attach.
DR InterPro: IPR000911; Ribosomal_L11.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr_1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHATANTHEINE; UNKNOWN_1.
DR PROSITE; PS00359; RIBOSOMAL_L11; UNKNOWN_1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 472 AA; 50203 MW; 4A0CBF6D60DB67E7 CRC64;

Query Match      1.5%; Score 8; DB 2; Length 472;

```

Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402
|||||||

Db 86 DRLGRKKT 93

RESULT 11

P71283 ID P71283 PRELIMINARY; PRT; 814 AA.
AC P71283;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL.
GN YAPF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U70214; AAB08643.1; -;
DR InterPro: IPR001552; Acyl-CoA_dh.
DR Pfam: PF00441; Acyl-CoA_dh; 1.
SQ SEQUENCE 814 AA; 89224 MW; AD9E40ACB44CF781 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 814;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 MWILSILA 108
|||||||

Db 1 MWILSILA 8

RESULT 12

P94078 ID P94078 PRELIMINARY; PRT; 1019 AA.
AC P94078; 096239;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ALPHA-MANNOSIDASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=97086699; PubMed=8932388;
RA Quigley F., Dao P., Cottet A., Mache R.;
RT "Sequence analysis of an 81 kb contig from Arabidopsis thaliana
RT chromosome III";
RL Nucleic Acids Res. 24:4313-4318(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Quigley F., Dao P., Cottet A., Mache R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; X98130; CAA66821.1; -;
DR EMBL; Y11767; CAA72432.1; -;
DR EMBL; AB026648; BAB01735.1; -;
DR Mendel; 17407; Arath; 2768; 17407
DR InterPro: IPR000602; Glyco_hydro_38.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF01074; Glyco_hydro_38; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
SQ SEQUENCE 1019 AA; 115219 MW; 4F764F2359F9FBC CRC64;

Query Match 1.5%; Score 8; DB 10; Length 1019;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LSGNQEKA 274
|||||||

Db 950 LSGNQEKA 957

RESULT 13

Q91703 ID Q91703 PRELIMINARY; PRT; 1060 AA.
AC Q91703;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FLII PROTEIN.
GN FLII OR BCDNA:LD21753 OR CGI484.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Aboyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:13 ; Search time 42.48 Seconds
(without alignments)
955.559 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 2846
Sequence: 1 MEEDLFQLRQLPVVKERRTG.....MVGKGMHGAGVTRSNSSGSQE 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database :
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 2846 | 100.0 | 548 | 21 | AAV44633 | Human organic cati |
| 2 | 493.5 | 17.3 | 439 | 22 | AAB76766 | Corynebacterium q1 |
| 3 | 493.5 | 17.3 | 448 | 22 | AAG93041 | C glutamicum prote |
| 4 | 441 | 15.5 | 535 | 21 | AAV51249 | Rat liver anion tr |
| 5 | 426.5 | 15.0 | 557 | 17 | AAB77676 | Rat OCT-1 protein. |
| 6 | 424 | 14.9 | 554 | 19 | AAW64538 | Human liver cell c |
| 7 | 402 | 14.1 | 542 | 22 | AAB47274 | HOAT3. Homo sapie |
| 8 | 402 | 14.1 | 742 | 21 | AAV94977 | Human secreted pro |
| 9 | 398 | 14.0 | 542 | 21 | AAV92902 | Human cerebral org |
| 10 | 396.5 | 13.9 | 536 | 21 | AAV92903 | Rat cerebral organ |
| 11 | 395 | 13.9 | 537 | 18 | AAW44195 | Mouse osteoclast t |

| | | | | | | |
|----|-------|------|-----|----|----------|--------------------|
| 12 | 392.5 | 13.8 | 540 | 22 | AAB9401 | Murine organic an |
| 13 | 387 | 13.6 | 561 | 18 | AAW41196 | Human osteoclast |
| 14 | 379 | 13.3 | 607 | 21 | AAW12131 | Hydrophobic dom |
| 15 | 374.5 | 13.2 | 550 | 21 | AAV44278 | hoAT1. Homo sapi |
| 16 | 374.5 | 13.2 | 550 | 22 | AAB47271 | Human organic an |
| 17 | 374.5 | 13.2 | 563 | 20 | AAW88489 | hoAT2B. Homo sapi |
| 18 | 370.5 | 13.0 | 538 | 22 | AAB47273 | hoAT2A. Homo sapi |
| 19 | 370 | 13.0 | 546 | 22 | AAB47272 | A human organic an |
| 20 | 367 | 12.9 | 548 | 21 | AAW08823 | Rat organic anion |
| 21 | 361.5 | 12.7 | 551 | 20 | AAW84488 | A human organic an |
| 22 | 360.5 | 12.7 | 551 | 21 | AAW08824 | A protein with cat |
| 23 | 357.5 | 12.6 | 557 | 20 | AAV01650 | Human carnitine tr |
| 24 | 357.5 | 12.6 | 557 | 21 | AAW83929 | Mouse organic anio |
| 25 | 348 | 12.2 | 545 | 22 | AAB36553 | A protein with cat |
| 26 | 332 | 11.7 | 557 | 20 | AAV01652 | Mouse OCTN2 amino |
| 27 | 332 | 11.7 | 557 | 21 | AAW63930 | Mouse carnitine tr |
| 28 | 332 | 11.7 | 557 | 21 | AAW63930 | Mouse OCTN3 protei |
| 29 | 329.5 | 11.6 | 564 | 21 | AAW20578 | C glutamicum prote |
| 30 | 316.5 | 11.1 | 431 | 22 | AAG90943 | Corynebacterium gl |
| 31 | 311.5 | 10.9 | 399 | 22 | AAB76717 | Cat flea HMT synap |
| 32 | 311.5 | 10.9 | 530 | 21 | AAB29626 | A protein with cat |
| 33 | 302.5 | 10.6 | 551 | 20 | AAV01649 | Human ORFX ORF2802 |
| 34 | 301.5 | 10.6 | 560 | 21 | AAB43038 | Corynebacterium gl |
| 35 | 297 | 10.4 | 446 | 22 | AAB76810 | C glutamicum prote |
| 36 | 297 | 10.4 | 475 | 22 | AAG92363 | Human protein havi |
| 37 | 296.5 | 10.4 | 550 | 22 | AAE06612 | Human organic anio |
| 38 | 296.5 | 10.4 | 584 | 22 | AAW00930 | Human bone marrow |
| 39 | 295 | 10.3 | 578 | 22 | AAE06571 | Human protein havi |
| 40 | 294.5 | 10.3 | 582 | 22 | AAE06571 | C glutamicum prote |
| 41 | 288.5 | 10.1 | 452 | 22 | AAG93094 | A protein with cat |
| 42 | 286.5 | 10.1 | 553 | 20 | AAV01651 | Mouse OCTN1 amino |
| 43 | 286.5 | 10.1 | 553 | 21 | AAW20579 | Plant phosphate tr |
| 44 | 284 | 10.0 | 585 | 22 | AAB47415 | C glutamicum prote |
| 45 | 283.5 | 10.0 | 460 | 22 | AAG92391 | |

RESULT 1

AAV44633

AAV44633 standard; Protein; 548 AA.

AC

AAV44633;

07-APR-2000 (first entry)

Human organic cation transporter-like protein (OCTlp).

Human; organic cation transporter-like protein; OCTlp; transporter; transmembrane; nootropic; neuroprotective; neuroleptic; anticonvulsant; antiParkinsonian; antidepressant; cellular process; cell proliferation; screen; treatment; prevention; diagnosis; neurodegenerative disorder; Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis; CNS disorder; central nervous system; schizophrenia; depression; behavioural; sleep disorder; Alzheimer's; eating disorder.

Homo sapiens.

OS

Key

Domain

Location/Qualifiers

1..85

/label= Cytoplasmic_domain

141..154

/label= Cytoplasmic_domain

199..208

/label= Cytoplasmic_domain

259..314

/label= Cytoplasmic_domain

395..402

/label= Cytoplasmic_domain

448..457

/label= Cytoplasmic_domain

511..548

/label= Cytoplasmic_domain

ALIGNMENTS

| | |
|----------|---|
| RESULT | 1 |
| RAY44633 | |
| ID | RAY44633 standard; Protein; 548 AA. |
| XX | |
| XX | RAY44633; |
| XX | |
| XX | 07-APR-2000 (first entry) |
| XX | |
| XX | Human organic cation transporter-like protein (OCTP); |
| DE | |

Human; organic cation transporter-like protein; OCT1p; transporter; transmembrane; norepinephrine; neuroprotective; neuroleptic; anticonvulsant; antiparkinsonian; antidepressant; cellular process; cell proliferation; screen; treatment; prevention; diagnosis; neurodegenerative disorder; Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis; CNS disorder; central nervous system; schizophrenia; depression; behavioural; sleep disorder; Alzheimer's; eating disorder.

| | | | |
|----|--|--|--|
| FT | | /label= Cytoplasmic_domain | |
| FT | Domain | 86..108 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 122..140 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 155..175 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 182..198 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 209..232 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 239..258 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 315..337 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 376..394 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 403..421 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 428..447 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 458..479 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 487..510 | |
| FT | | /label= Transmembrane_domain | |
| FT | Domain | 109..121 | |
| FT | | /label= Extracellular_domain | |
| FT | | /note= "Hydrophilic region useful for antibody production" | |
| FT | Domain | 176..181 | |
| FT | | /label= Extracellular_domain | |
| FT | | /note= "Hydrophilic region useful for antibody production" | |
| FT | Domain | 233..238 | |
| FT | | /label= Extracellular_domain | |
| FT | | /note= "Hydrophilic region useful for antibody production" | |
| FT | Domain | 338..375 | |
| FT | | /label= Extracellular_domain | |
| FT | | /note= "Hydrophilic region useful for antibody production" | |
| FT | Domain | 422..427 | |
| FT | | /label= Extracellular_domain | |
| FT | | /note= "Hydrophilic region useful for antibody production" | |
| FT | Domain | 480..486 | |
| FT | | /label= Extracellular_domain | |
| FT | | /note= "Hydrophilic region useful for antibody production" | |
| FT | Region | 71..524 | |
| FT | | /note= "shows homology to a consensus sequence for sugar and other transporter molecules derived from a hidden Markov model" | |
| XX | | | |
| PN | WO200000633-A1. | | |
| XX | | | |
| PD | 06-JAN-2000. | | |
| XX | | | |
| PF | 29-JUN-1999; | 99WO-US14880. | |
| XX | | | |
| PR | 30-JUN-1998; | 98US-0107932. | |
| XX | | | |
| PA | (MILL-) MILLENNIUM PHARM INC. | | |
| XX | Goodearl AJ, Glucksmann MA; | | |
| PI | | | |
| XX | | | |
| DR | WPI: 2000-137069/12. | | |
| NR | -PSDB; AA249886. | | |
| XX | | | |
| PT | New nucleic acid encoding human organic cation transporter-like protein, used for prevention, treatment and diagnosis of e.g. neurological, behavioural or sleep disorders - | | |
| XX | | | |

PS
xx
CC The present sequence is a human OCTIP (organic cation
CC transporter-like protein), a member of the superfamily of sugar and
CC other transporter molecules that have 12 transmembrane domains.
CC The sequence is derived from a human foetal brain cDNA library. The
CC protein is highly expressed in brain tissue and has nootropic,
CC neuroprotective, neuroleptic, anticonvulsant, antiParkinsonian,
CC antidepressant activities. The present sequence is used to
CC regulate a variety of cellular processes e.g. cell proliferation,
CC differentiation and survival, screen OCTIP modulators and detect mutation
CC in OCTIP gene. OCTIP modulators can be used to treat or prevent chronic
CC neurodegenerative disorders (e.g. Alzheimer's, Parkinson s,
CC Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic,
CC depression), behavioural, sleep and eating disorders.

xx
S0 Sequence 548 AA:


```

RESULT 5
AAR77676
ID AAR77676 standard; Protein: 556 AA.
XX
AC AAR77676;
XX
10-SEP-1996 (first entry)
XX
DE Rat OCT-1 protein.
XX
KW Rat; OCT-1; transporter protein; cationic; xenobiotic; pharmaceutical;
KW blood; liver; kidney; epithelial cell; intestine; tetraethylammonium;
KW proximal renal tubule cell; intestine; enterocyte; transgenic; renal;
KW biliary; excretion; resorption; modulator; uptake.
XX
OS Rattus rattus.
XX
FH Key Location/Qualifiers
FT Domain 20..46 /note= "transmembrane domain"
FT Modified-site 71 /note= "putative N-glycosylation site"
FT Modified-site 97 /note= "putative N-glycosylation site"
FT Modified-site 113 /note= "putative N-glycosylation site"
FT Domain 154..171 /note= "transmembrane domain"
FT Domain 178..197 /note= "transmembrane domain"
FT Domain 243..260 /note= "transmembrane domain"
FT Domain 267..283 /note= "transmembrane domain"
FT Domain 350..366 /note= "transmembrane domain"
FT Domain 380..398 /note= "transmembrane domain"
FT Domain 406..425 /note= "transmembrane domain"
FT Modified-site 432 /note= "putative N-glycosylation site"
FT Domain 435..452 /note= "transmembrane domain"
FT Domain 469..485 /note= "transmembrane domain"
FT Domain 494..514 /note= "transmembrane domain"
XX DE424577-A1.
XX
XX 18-JAN-1996.
XX
XX 13-JUL-1994; 94DE-4424577.
XX
XX 13-JUL-1994; 94DE-4424577.
XX
XX (FARH ) HOECHST AG.
XX
XX Gorboulev V, Gruendeman D, Koepsell H;
XX
XX WPI; 1996-069422/08.
XX
XX N-PSDB; AAT08702.
XX
XX Transporter protein for cationic xenobiotic(s) and pharmaceuticals,
XX and related DNA and transformed cells - used e.g. to assess
XX excretion and resorption of cationic cpds.
XX
XX Claim 1; Fig 2A1; 13pp; German.
XX
XX This is the amino acid sequence of the rat OCT-1 protein, a new
CC

```

```

CC transporter protein able to transport cationic xenobiotics and
CC pharmaceuticals from the blood into liver or kidney epithelial
CC cells or from the intestine. The gene was isolated by injecting
CC a rat kidney gene library into Xenopus laevis oocyte and isolating
CC clones conferring uptake of 14C-tetraethylammonium. One clone
CC designated OCT-1 was isolated. Expression of the gene was detect
CC in proximal renal tubule cells, in liver epithelial cell and in
CC intestinal enterocytes. The DNA can be used to generate transgenic
CC cells for use in in vitro test for renal/biliary excretion or
CC intestinal resorption of xenobiotics and pharmaceuticals. The protein
CC or cells expressing it can also be used to isolate modulators that
CC block uptake of pharmaceutical by the renal tubules.
XX
XX Sequence 556 AA;
SQ
Query Match 15.0%; Score 426.5; DB 17; Length 556;
Best Local Similarity 29.3%; Pred. No. 2.2e-38;
Matches 127; Conservative 75; Mismatches 167; Indels 65; Gaps 12;
QY 119 SQVALLTSVVFVGMMSSSTLWGNISDQYGRKTKISVLWTLXYGILSAFAPVYSWILV 178
DB :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 awkvdllfqcenvnlgfllgslvvgyladrfgkclclvltlvtlvtavapdytsml 205
QY 179 LRGIVGF-GIGGVPOSVTLYAEFLPMKARAKCILLIEVFAICTVFEVVLAVFVMPSLGW 237
DB |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 frllqgmvsksvsgvgtllitefvgsgyrrttaillyqmaftvgivglagva-yaipd--w 262
QY 238 RWLLILSAVPIILLFAVLFCFWLPESARYDVLSGNQEKAATATLKRIATENGAMPPLGKLIIIS 297
DB ||| :||| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
263 rwlqlavslptflilywfvpsprwlisqktrtravimeqlaqngkvyppadlknlc 322
QY 298 RQDRGKMR-----DLF-TPHFRWTTLLLPWFNSAFSYVGLVLTTELFGAGDVCGIS 351
DB ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 leedasekrspsadlrrtptlrkhtvlmylwfscavlyqgllm----- 367
QY 352 SRKKAWEAKCSLACEYLSSEEDYMDLLTTLSEFPGLVLTWLIDRLGRKKTWALC-FVIF 410
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 -hvgatganl-----yldffyslvfepaafllvldrlgrfyiaasnlvtg 415
QY 411 SFCSLLLFI-----CVGRNVLTLILLFIARAFISGGFOAAVYVTPVPTATRA 458
DB :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 aacllmifiphelhwlnvtlaciqrngativl-----qmvclvnaelyptflrn 464
QY 459 IGLGTCGSMARVAGALITPFIQAQVWLESVVLTAVYSGCCLLAALASCFLPIETKGGLQ 518
DB ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 lgmvmvscaldlgiftfpmvfrlmevwqalpdlilfgvlgtagamtllip-etkqv alp 523
QY 519 ESSHREWGQEMVGR 532
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
524 etlee---aenlgr 534
RESULT 6
AAW64538
ID AAW64538 standard; Protein: 554 AA.
XX
AC AAW64538;
XX
XX 21-OCT-1998 (first entry)
XX
XX Human liver cell clone HP01293 protein.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
XX Homo sapiens.
XX
XX WO9821328-A2.
XX
XX 22-MAY-1998.

```

```
XX 07-NOV-1997; 97WO-JP04056.
XX
XX 13-NOV-1996; 96JP-0301429.
XX
XX (PROT-) PROTEGENE INC.
XX (SAGA) SAGAMI CHEM RES CENTRE.
XX
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
XX WPI; 1998-297932/26.
XX N-PSDB; AAV49558, AAV49559.
XX
XX Human protein having transmembrane domain - useful for, e.g.
XX research and nutrition
XX
XX Claim 1; Page 96-98; 205pp; English.
XX
XX AAM64534-W64558 represent human proteins containing a transmembrane
XX domain. These proteins can be used for, e.g. research and nutrition, and
XX may have cytokine and cell proliferation/differentiation, immune
XX stimulating/suppressing, haematopoiesis regulating, tissue growth,
XX activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
XX receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
XX Sequence 554 AA:

Query Match 14.9%; Score 424; DB 19; Length 554;
Best Local Similarity 29.9%; Pred. No. 4.le-38;
Matches 126; Conservative 69; Mismatches 165; Indels 62; Gaps 11;

Qy 119 SNOVALLSVVFGVGMSSSTLWGNISDOYGRKTKLISVLTWLYYGLTSAFA.VYSWILV 178
Dy 145 swklldfgscinagffgslgvgfyfadrfgkcllgtvlvnavsgvlmafspnysmll 204
Qy 179 LRGLVGF-GIGGVQSVTLVLAETFLPMKARAKCILLIEVFVFAIGTVFVVLAVFVMPSLGW 237
Dy 205 frllqglvskgnwmagytllitefvgsgsrvtvaamyqmaftvgvlvltgla-yalph--w 261
Qy 238 RWLLILSAVPLLEAVLCFVLPESARYDVLSCNOEKALATLKRIATENGAPMLGKLIIIS 297
Dy 262 rwlqlavslptflllywvcpesprwllsqkrnteakimdhiaqknkkipadlknls 321
Qy 298 ROED-----RQKMRDLF-TPHFRWTTLLLPWFNSAPSYVGLVLLTTELFQAGDVCGIS 351
Dy 322 leedvteklspsfadlrtprrtkrftlmylwtfdsvlygglil-----hmgatsg-- 373
Qy 352 SRKRAVEAKCSLACEYLSEEDYMDLLTTLSEFPVGLVTLWIDRLGRKKTWALCFVIF- 410
Dy 374 -----nlylflsalvpeipgafialitdrvgrypmavsnllag 414
Qy 411 SFCSLLLFI-----CVGRNVLTLLLTARAFISGFOAAVYVTPPEVYPTAFRA 458
Dy 415 aacivmfispldbwlniimcvgrmgitai-----qmiclvnaelyptfvrn 463
Qy 459 LGLGTCSGMARVALITPFAQVNMLESVYTLAVYSCCLLAALASCFLPIETKGGGLQ 518
Dy 464 lgvnmvcsliodiggitpfvfrrevwqalplilfavlgilaagvtlllp-etkqvalp 522
Qy 519 ES 520
Dy 523 et 524

RESULT 7
AAB47274
ID AAB47274 standard; Protein; 542 AA.
XX
XX AAB47274;
AC
XX DT 06-AUG-2001 (first entry)
XX
```

```
DE HOAT3.
XX
XX Human; organic anion transporter; hoAT; liver; kidney;
XX membrane protein; transport; organic anion; splice variant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 259
XX /note= "PKC phosphorylation site"
XX Modified-site 266
XX /note= "PKC phosphorylation site"
XX Modified-site 269
XX /note= "PKC phosphorylation site"
XX Modified-site 511
XX /note= "PKC phosphorylation site"
XX Modified-site 527
XX /note= "PKC phosphorylation site"
XX
XX WO200104283-A2.
XX
XX 18-JAN-2001.
XX
XX 12-JUL-2000; 2000WO-US18980.
XX
XX 12-JUL-1999; 99US-0143771.
XX
XX (META-) METABASIS THERAPEUTICS INC.
XX
XX Sun W;
XX
XX WPI; 2001-367057/38.
XX N-PSDB; AAC85822.
XX
XX Nucleic acids encoding human organic anion transporter polypeptides,
XX useful in gene therapy procedures -
XX
XX Claim 2; Fig 4; 95pp; English.
XX
XX The sequences given in AAB47271-76 represent human organic anion
XX transporter (hoAT) polypeptides. hoAT polypeptides are preferentially
XX expressed in the liver and kidneys of humans. hoAT's are membrane
XX proteins that facilitate the transport of organic anions across the
XX cell membrane. The mechanism of transport is thought to be a secondary
XX or tertiary active transport involving exchange of another organic anion.
XX hoAT2A and hoAT2B are thought to be splice variants as they are
XX identical except at the C-terminal end. hoAT proteins and the DNA
XX encoding them, may be used in the prevention, treatment and diagnosis
XX of diseases associated with inappropriate hoAT expression.
XX
XX Sequence 542 AA;

Query Match 14.1%; Score 402; DB 22; Length 542;
Best Local Similarity 24.0%; Pred. No. 1.le-35;
Matches 131; Conservative 81; Mismatches 190; Indels 144; Gaps 17;

Qy 72 DAVEAIG-FGRFQWKLSTLTGLAW--MADAMEMMTLSILAPQLHCE-----WRLP- 118
Dy 5 eildrvsgmghfqlhvailgilnmanhnllqiftaatpvhhcrpphnastgppwlp 64
Qy 119 -----SNOVA-----L 124
Dy 65 gpngkperclrfvhhpnaslpndtqrnecldgwwynstkdsivtewdlvcnsnklkem 124
Qy 125 LTVSVFVGMSSSTLWGNISDOYGRKTKLISVLTWLYYGLTSAFAPVYVSTLVLRLGLVG 184
Dy 125 aqsfimgiliggivlgdlsrfrprrpiltcsylllaasgsgaafspfiymvfrfcg 184
Qy 185 FGIGGVQPS-VTLVLAETFLPMKARAKCILLIEVFVFAIGTVFVVLAVFVMPSLG-----WR 238
Dy 185 fgisgitltilnvewvptmraimstalgycytfqg-----filpglayalpqr 236
```

QY 239 WLLILSNVPLLLFAVLCFWLPESARYDVLGNOEKAIATLKRIAT-----ENGAPMPLCK 293
Db 237 wlgitvsiptvflfsswwtpeisirvlvisgksskalkilrrvavfngkkeegeerlslee 296
QY 294 LITSROED-----RGKMRDLF-TPHFRWTTLLWFIWFSNAFSYXGLVLLTTE----- 340
Db 297 lkinlqkeislakakytasdlfrpmlrmtfclsawfatgfayyslangveefgvnly 356
QY 341 -----LFOAGDVUGISSRKKAVEAKCSLACEYLSEEDYMDLLWTLSEFPGLVTLWIHDR 396
Db 357 ilqifggvdv-----pakfittilsy 379
QY 397 LGRKKTALCFVIFSCSL-LLFICVG-RNVLTLLFIARAFISGGFOAAAYVTPEVYPT 454
Db 380 lgrhtqaallagaaialtvtpldqlcvrtvavfkgkclsssfcliytsetlypt 439
QY 455 ATRALGIGTSGMARVAGALITPPIAOVMLESSVYLLTAVYSGCLLAALASCFLPIETKG 514
Db 440 vtrqtmgvsnlwtvrgsmvslv-kitgevqfpipniygitalligssaalfip-etln 497
QY 515 GGLQES 520
Db 498 qplpet 503
RESULT 8
AA94977
ID - AA94977 standard; Protein: 742 AA.
AC AA94977;
XX
XX 16-JUN-2000 (first entry)
XX Human secreted protein clone as180_1 protein sequence SEQ ID NO:160.
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy.
XX Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX (GBMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy MJ, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPT; 2000-205979/18.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 169; Page 614-616; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94980 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta. The polynucleotides and proteins are
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted protein; from the present invention.
XX
XX Sequence 742 AA;
QY 21 ESARSEDDTASGE-----VOIEGVHVGLEAVEIDNDGAAP----- 57
Db 85 eghdedddiyegeyqgpraesggkgermadgaplagvrg-----lsdgcpgpggrea 139
QY 58 -----KEFANPTDDTFWEDAVEAIGFQWKLVSVLTLGLANWADAMEMMILSILAPQL 111
Db 140 qrtkereeiaq-----gyeailrecqhgfrqwtlyfvlglalmadgvevfvvqvlpsa 193
QY 112 HCEWRLFSQWVALLTSVVFVGMSSSTLWGNISDQYGRKTKISVLTWLYGILSAFAP 171
Db 194 ekdmclsdsnkgmlglivymvgafwgggladrigrqcllslsvnsvfaiffsvq 253
QY 172 VYSWILVLRGLVGFIGG-VPOSVTLTAEFLPMKARAKCILLIEVFWAIGTVFEVWLAVF 230
Db 254 gygtflcrllsgvgigspivfsyfeqlaqekrgchlswlcmfwmiggyaamaawa 313
QY 231 VMPSLG-----WRWLLILSAVPLLLFAVLCFWLPESARYDVLGNOEKAIAT 277
Db 314 ilphygwsfmgssayqhsrwrvfivcappsvfaigalttqpesprfflengkhdeamv 373
QY 278 LKRIATEN-----GAP---MPLGKLIISRQEDR-----G 303
Db 374 lkqvhdntmrakghpervsvthiktihqedellieqsdgtgtwygrvralsgggvq 433
QY 304 KMRDLTFPHFRWTTLLWFIWFSNAFSYXGLVL----- 336
Db 434 nlfscfgpeyrritlmumvgvwtmsfyygltvwpdmirhlqavdyasrtkvfpgerve 493
QY 337 -----LTTELFOAGDVCG-----ISSRKKAVEAKCSL-----ACEYLS 369
Db 494 hvtfntlenqhrqggqfndkfgrlksvsvfedsifeecyfedvtssntffnctfin 553
QY 370 EEDY-----MDLLWT-----TLSEFPGLV 388
Db 554 tvfyntdlfeykfvnsrlinstflhnkegcpldvtgtgegaymyvsvflgtlavlpgni 613

Query Match 14.1%; Score 402; DB 21; Length 742;
Best Local Similarity 21.8%; Pred. No. 1.8e-35;
Matches 148; Conservative 99; Mismatches 231; Indels 200; Gaps 19;

| | | | |
|---------------------------|--|--|-----|
| QY | 389 | VTLWIIIDRLGRKKTALCFVIFSCGSLLLFICVGRN---VLTKLLFIARAFISGGFOAAAY | 445 |
| Db | 614 | vsallmdkigrlrmlagssvm--scvsofflsgnseamiallcifggvsia-swnald | 670 |
| QY | 446 | VVTPEVYPPTATRALGLTGCSGARVGALI-----TPFIAQVMLESSYYTTLAVYSGCCLL | 500 |
| Db | 671 | vltvelypsdkrttafgfinaicklaavligisftsvgitkaapilfasaalalgslla | 730 |
| QY | 501 | AALASCFLEPIETKGSGGIQ 518 | |
| Db | 731 | lklp-----etrgqvlq 742 | |
| RESULT | 9 | | |
| AAAY92902 | | | |
| ID | AAAY92902 | standard; Protein; 542 AA. | |
| AC | AAAY92902; | | |
| XX | 26-SEP-2000 | (first entry) | |
| XX | Human cerebral organic anion transporter OAT3 protein. | | |
| XX | Human; organic anion transporter protein; OAT3; cerebral tissue. | | |
| OS | Homo sapiens. | | |
| XX | WO200017237-A1. | | |
| PN | 30-MAR-2000. | | |
| PD | 20-SEP-1999; | 99WO-JP05120. | |
| PF | 18-SEP-1998; | 98JP-0265126. | |
| PR | (NISC-) JAPAN SCI & TECHNOLOGY CORP. | | |
| PA | Endou H, Sekine T, Kusuhsara H; | | |
| XX | WPI; 2000-283546/24. | | |
| DR | N-PSDB; AAA11146. | | |
| XX | Organic anion transporter protein OAT3 expressed in cerebral tissue, | | |
| PT | antibodies to it and gene encoding it for study of cerebral transport | | |
| PT | of anions including drug molecules - | | |
| XX | Claim 2; Page 18-20; 48pp; Japanese. | | |
| PS | This sequence represents a human organic anion transporter protein OAT3. | | |
| CC | The sequence is used to regulate the transport of anions in cerebral | | |
| CC | tissue, including drug molecules, and regulation of this transport. | | |
| XX | Sequence 542 AA; | | |
| SQ | | | |
| Query Match | 14.0%; Score 398; DB 21; Length 542; | | |
| Best Local Similarity | 23.8%; Pred. No. 3.2e-35; | | |
| Matches 130; Conservative | 82; Mismatches 190; Indels 144; Gaps | | |
| 17; | | | |
| QY | 72 | DAVEAIG-FGKFOWKLSVLTLGLAW--MADAMEMMLISTLAPQLHCE-----WRLP- | 118 |
| Db | 5 | eildrvsgmhgfqlhvailglplmanhnllqiftaatpvhhcrphpnastgwpvlpm 64 | |
| QY | 119 | -----SQVA-----L 124 | |
| Db | 65 | gpngkperclrfvhppnasipndtqramepcldgwvynstksdvitewdlvcnsnkikem 124 | |
| QY | 125 | LTSYVFVGMMSSTPLWGNISDOYGRKTGLKISLVLTLYYGILSAFAPVYSWILVRGLVG 184 | |
| Db | 125 | aqsifmagiliglvigdlsdrgrpilctsylllaasgsgaaftsfpjymvfrflcg 184 | |
| QY | 185 | FGIGVPQS-VTLYAELFLPKAKACTILLIEVFAIGTVFEVVLAVFMPSLG----WR 238 | |

[illegible]

Query Match 13.9%; Score 396.5; DB 21; Length 536;
Best Local Similarity 28.3%; Pred. NO. 4.6e-35;
Matches 119: Conservative 79; Mismatches 173; Indels 49; Gaps 15;

| Query Match | 13.9%; | Score 395; | DB 18; | Length 537; |
|-----------------------|------------------|--|-------------|-------------|
| Best Local Similarity | 23.6%; | Pred. No. 6.8e-35; | | |
| Matches 130; | Conservative 91; | Mismatches 194; | Indels 136; | Gaps |
| Qy | 72 | DAVEAIG-FCKFWKLSVLTGLAWMADAME--NMILSILAPQLHCE-----WRLP-118 | | |
| Db | 5 | eildrvsgmpdyhvtllalpilgianhmlqifattpdhcrpppnaslepwlpl 64 | | |
| Qy | 119 | -----SWQVA-----L 124 | | |
| Db | 65 | gpngkpekclrfvhlpnasipndtqgatepcldgwginystrdtivtcwdlvogsnklkem 124 | | |
| Qy | 125 | LTSVVFVGMSSSTLGNISDQYGRKTGLKISVLWLTLYGILSAFAPVSWILVLRGLVG 184 | | |
| Db | 125 | aqsvmagilvgpvfgelsdrfgrkpiltwsyllaagssaaftspsitvymifrlcg 184 | | |
| Qy | 185 | FGIGGVPOS-VTVIAEFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVFWPMSLG-----WR 238 | | |
| Db | 185 | csisgislstliinvewpplstraissttligcyctlqg-----filpglayavpqr 236 | | |
| Qy | 239 | WLILSVAPLLLFAVICFWLPESARYDVUSGNOEKAIATLKRIAT-----ENGAPMPLCK 293 | | |
| Db | 237 | wlqlsvsaafifsliswwypesirwlvlsngkfsralktlgrvatfngkkeegekitvee 296 | | |
| Qy | 294 | LILSRQED-----RGKMRDLF-TPHFRWTLLLMFWFSNAPSYGVLVLLTTTELFAQGD 346 | | |
| Db | 297 | lknlgkdtisakvkvgisdrfvsrlrvfclslawfalgfayyslmgveef-----351 | | |
| Qy | 347 | VCGTSSRRKKAWEAKCSLACVSEEDYMDLLWTTLSEFFGVLVTLMIIDIRGRKKTMAIC 406 | | |
| Db | 352 | --gvni-----yl-----lqifggv-dipakftilsisylgritqgfl 389 | | |
| Qy | 407 | FVIFSFCSL-LLFICVGRNVL-TLLLFIFARAFISGGFQAAYVYTPVYPTATRAIGLGTG 464 | | |
| Db | 390 | liitagvaialifvssemqilrtalavfgkclsgsfscfllytselyptvirqtgmgis 449 | | |
| Qy | 465 | SGMARVCALLTPPTI-----AQMLESVVLTUAVYSGCCLLAALASCFPIETKGGGLOE 519 | | |
| Db | 450 | niwarvmslapivkkitgelqpfpnvifwtmllgg-----saaffiletlnrlpe 502 | | |
| Qy | 520 | SSH--REWGQE 528 | | |
| Db | 503 | tiediaqdwaaq 513 | | |

| | |
|-----------|---|
| RESULT 12 | |
| AAB49401 | |
| ID | AAB49401 standard; Protein; 540 AA. |
| XX | |
| AC | AAB49401; |
| XX | |
| DT | 02-MAR-2001 (first entry) |
| XX | |
| DE | Murine organic anion transporter 6. |
| XX | |
| KW | Murine; organic anion transporter 6; mOATP6; cancer; inflammation; |
| KW | cardiovascular disease; central nervous system disorder; kidney disease |
| KW | liver disease; autoimmune disease. |
| XX | |
| OS | Mus sp. |
| XX | |
| PN | WO200070048-A1. |
| XX | |
| PD | 23-NOV-2000. |
| XX | |
| PF | 15-MAY-2000; 2000WO-US13316. |
| XX | |
| PR | 14-MAY-1999; 99US-0134137. |
| PR | 12-MAY-2000; 2000US-0570293. |
| XX | |
| PA | (SMIK) SMITHKLINE BEECHAM CORP. |
| PA | (SMIK) SMITHKLINE BEECHAM PLC. |

Db 380 lgrhttaqaalllraggailaltftvpidlqtvtvavfkgkclssstscfllytselypt 439

QY 455 ATRALGLGTCGSMARVCALITPEIAQVMLESSVYLTLAVYSGCLLAALASCFLEPIETKG 514

Db 440 vtrgtmgvsnlwtvgmsvplv-kitgevqpfipnilygitaliggsaalflp-ettln 497

QY 515 GGLOES 520

Db 498 qp1pet 503

RESULT 14

AAB12131

ID AAB12131 standard; Protein: 607 AA.

AC AAB12131;

XX 02-FEB-2001 (first entry)

XX Hydrophobic domain protein from clone HP03092 isolated from Liver cells.

XX Human; secreted protein; membrane protein; hydrophobic domain;

KW proliferation control; differentiation induction; material transport;

KW biophylaxis; signal receptor; ion channel; transporter; immunostimulant;

KW immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;

KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;

KW autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer.

XX Homo sapiens.

XX W0200029448-A2.

XX 25-MAY-2000.

XX 17-NOV-1999; 99WO-JP06412.

PR 17-NOV-1998; 98JP-0326255.

PR 22-DEC-1998; 98JP-0364315.

PR 16-MAR-1999; 99JP-0069811.

PR 27-APR-1999; 99JP-0119299.

PR 19-MAY-1999; 99JP-0138169.

XX (SAGA) SAGAMI CHEM RES CENT.

PA (PROT-) PROTEGENE INC.

XX Kato S, Kimura T;

XX WPI; 2000-387753/33.

DR N-PSDB; AAA62004, AAA62014.

XX Proteins comprising hydrophobic regions, such as secretory and membrane

PT proteins, useful in research and diagnostics and having various

PT activities e.g. immunomodulatory, antiinflammatory, chemokinetic,

XX hemostatic, thrombolytic -

Claim 1: Page 236-238; 410pp; English.

XX Secretory proteins play important roles in the proliferation control, the

CC differentiation induction, the material transport and the biophylaxis of

CC cells. Membrane proteins have important roles as signal receptors, ion

CC channels and transporters. The present sequence is a human protein which

CC has at least one hydrophobic domain. This protein may be a secretory or a

CC membrane protein. The present protein may have cytokine and cell

CC proliferation/differentiation activity, immune stimulating or suppressing

CC activity, haematopoiesis activity, tissue growth activity,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, anti-inflammatory activity and tumour

CC inhibition activity. The present protein could therefore be used for

CC treatment of autoimmune disease, Alzheimer's disease, Parkinson's

XX disease, and cancer.

XX Sequence 607 AA;

Query Match 13.3%; Score 379; DB 21; Length 607;

Best Local Similarity 26.5%; Pred. NO. 5e-33;

Matches 140; Conservative 76; Mismatches 204; Indels 108; Gaps 18;

QY 29 TASGEHEVQIEGVHVGLEAVELDD-----CAAVPKFANPTDDTFVDEDAVEALGFGKEQ 83

Db 127 tiatesqvgviyhlevecrwrspweaagrglpwea-----eaaglgdrk 173

QY 84 -----WKLS---VLTGLAMMADAMEMMILSILAPOLHCEWRLPSSQVAL---LTSVVFV 131

Db 174 vsyspswreslgglsgmew-----dlvce-----qklnraastffa 212

QY 132 GMMSSSTLGNISDQGRKTKLISVLWTLTYGILSAFAPVYSWILVLRGLVGFIGGVP 191

Db 213 gvlvgavafgylsdrfgrtrlllvayvstlvglaasvymfalftrltgsalagft 272

QY 192 QSV-TLYAEFLPMKAKAKICILLIEVFMAIGTVEFVLAIVFVMPSLGMRWLLILLSAVPILL 250

Db 273 ilvmplelewdvehr:vagvlsstftwtggymllaivgyllr---dwrwillavtlpcap 329

QY 251 FAVLCWLPESARYDVLGNGQEKAIATLKRIATENGAP-----MPLGKLIIIS 297

Db 330 gillslwvpesarwlltqghvkeahryllhcarlngprvcdedsfsqeavskvaagerivr 389

QY 298 QEDRGCKMRDLF-TPHFRWTTLLWFIFSNFASYGLVLLTTELFOAGDVCGISRRKKA 356

Db 390 rp-----syldftrprlrlhslccvvvvgvnfsyyglsel-----dvsg----- 429

QY 357 VEAKCSLACEYLSSEEDYMDLLTTLSEFPGLVLTWIIDRLGRKKTKMA--LCFVIFSFC 414

Db 430 -----lglnvyqtqlfgavelpskllvylsvryagrrltqagtlgtalagft 478

QY 415 LLLFICVGRNVLTLLLFIAFAFISGGFOAAVYVTPPEYPTATRALGLGTCGSMARVCALI 474

Db 479 rllvssdmkswtvavmgkafseaafttaylftselvptvtrtgmglcalvgrl99sl 538

QY 475 TPFTAQVMLESSVYVLT--AVYSGCCLLAALASCFLEPIETKGGGLQES 520

Db 539 aplaa---llldgwlsipkityggiallaagtalilp-etrqqlpet 582

RESULT 15

AA44278

ID AAY44278 standard; Protein: 550 AA.

XX AC AAY44278;

XX DT 29-FEB-2000 (first entry)

XX Human organic anion transporter.

XX Human organic anion transporter; hOAT; nephrotoxic compound; screen;

KW drug-drug interaction; nucleotide phosphate.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 515..528

FT /label= Immunogen

FT /note= "For producing antibodies"

XX W09964459-A2.

XX 16-DEC-1999.

XX 10-JUN-1999; 99WO-US13172.

XX 11-JUN-1998; 98US-0088864.

PR 03-MAY-1999; 99US-0132267.

XX (GILE-) GILEAD SCI INC.

```

PI Cihlar T;
XX WPI: 2000-097519/08.
XX N-PSDB; AAZ29300.
XX
PT Human organic anion transporter nucleic acid used in drug screening,
PT identifying drug-drug interactions or individuals in whom particular
PT drugs are nephrotoxic -
XX
XX Claim 9; Fig 1; 36pp; English.
XX
CC The present sequence is human organic anion transporter (hoAT) protein.
CC hoAT is expressed in kidney and brain and removes toxic anions
CC from the circulation. It can be produced in host cells by transforming
CC them with recombinant vectors containing nucleic acid encoding hoAT.
CC It can be used in screening for compounds that suppress or enhance anion
CC uptake and transport by hoAT, to identify molecular variants of
CC nephrotoxic compounds, to detect drug-drug interaction within kidney
CC and brain and to test for transport of a covalently modified form of a
CC nucleotide phosphonate analogue by hoAT.
XX Sequence 550 AA;
XX
Query Match 13.2%; Score 374.5; DB 21; Length 550;
Best Local Similarity 25.6%; Pred. No. 1.4e-32;
Matches 132; Conservative 86; Mismatches 207; Indels 91; Gaps 18;
QY 54 AAVPKETANPTDDTFMVEDA-----VEAIGFKFQWKLSVLTKL----- 92
| | | : : : : : | | | | | | | | |
Db 42 aaaihhcrppadanlsknsglewlprdrqgqpescifrtspwgplfngteangtga 101
QY 93 -----AMMADAMEMMILSIAPQLHCEWRLPMSQVA---LLTSVFVGMSSSLTWNLI 143
| | | : : : : : | | | | | | | | |
Db 102 tepctdgiydn-----stfpstivtewdlvcshralrqlaqslymgvllgamvfyl 155
QY 144 SDQYGRKTGLKISLVLTLYYGILSAFAPVYSWILVRGLVGFGICGVP--QSVTLYAEFLP 202
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 156 adigrtrkvlilnylqtavscacafapfnpiycatfrllsgmalagisincmtlnvemwp 215
QY 203 MKAKACILLIEVFAIWGTVEVVLAIVPMPSLGMRWLIIISAVPLLIFAVLCFWLPESA 262
| | | : : : : : | | | : : : : : | | | : : : : : | | |
Db 216 iht racvqtlligyyslqg-fllagvayavph-wrhqllyvsapffaffiysswfiesea 272
QY 263 RYDLVSGNQEKAIATLKRIA-----TENGA MP LCKLIISRQED--RKG-----MRDLP 310
| | | : : : : : | | | : : : : : | | | : : : : : | | |
Db 273 rwhssgtldltalrlqrvaringkreagaklsmeivrasiqelmtgkgkasameilir 332
QY 311 PHFRWTTLTLWFIFNSAFSYGYGVLLTTTELFOAGDVCGISSRRKAWEAKSLACEYLSE 370
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 333 ptlrhlfcmlwlfatsfyglvm-----dlggfg-----vsiyliq 371
QY 371 EDY--MDLLTWLTLSFPQGVLTMIIDRLGRKKTNMALCFIPFSCLLLLFCVGNRN--VL 426
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 372 vifgavl-----paklvgfivinslgrtpadmaalltagiclingvipqdgsivr 423
QY 427 TLLLLFIARAFISGGFOAAVYTPEYPPTATRALGI TGTCGSMARVGALITPFPTAQVMLESS 486
| | | : : : : : | | | : | | : | | : | | : | | : | | : | |
Db 424 tsvlavlgkcilaasfnclifytgelytmirtqmgmstmarmarvgisvpivs----mtae 480
QY 487 VY-LTVAYSGCCLLAALASCFPLPIETKGGGLQES 520
Db 481 lypsmplfiygavpvvaasavtlvip-etlqqplpdtt 515

```

Search completed: March 13, 2002, 12:41:06
Job time: 53 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 21.65 seconds
(without alignments)
569.598 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 2846
Sequence: 1 NEEDLFQLRLPVVKERRTG.....MVGRMHGAGVTRNSGSQE 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------|
| 1 | 429 | 15.1 | 555 | 3 | US-08-501-572-3 |
| 2 | 429 | 15.1 | 555 | 3 | US-09-040-444-3 |
| 3 | 426.5 | 15.0 | 556 | 3 | US-08-501-572-1 |
| 4 | 426.5 | 15.0 | 556 | 3 | US-09-040-444-1 |
| 5 | 425.5 | 15.0 | 553 | 3 | US-08-501-572-2 |
| 6 | 425.5 | 15.0 | 553 | 3 | US-09-040-444-2 |
| 7 | 395 | 13.9 | 537 | 2 | US-08-647-397-2 |
| 8 | 217 | 7.6 | 584 | 2 | US-08-928-692-13 |
| 9 | 203.5 | 7.2 | 520 | 4 | US-08-968-127-2 |
| 10 | 203.5 | 7.2 | 524 | 2 | US-08-928-692-12 |
| 11 | 178 | 6.3 | 493 | 2 | US-09-031-392-10 |
| 12 | 178 | 6.3 | 493 | 4 | US-09-299-549-10 |
| 13 | 171 | 6.0 | 494 | 2 | US-09-031-392-5 |
| 14 | 171 | 6.0 | 494 | 4 | US-09-299-549-5 |
| 15 | 157 | 5.5 | 488 | 2 | US-08-928-692-11 |
| 16 | 154.5 | 5.4 | 500 | 2 | US-09-031-392-7 |
| 17 | 154.5 | 5.4 | 500 | 4 | US-09-299-549-7 |
| 18 | 154.5 | 5.4 | 509 | 2 | US-09-031-392-6 |
| 19 | 154.5 | 5.4 | 509 | 4 | US-09-299-549-6 |
| 20 | 150.5 | 5.3 | 534 | 2 | US-09-031-392-4 |
| 21 | 150.5 | 5.3 | 534 | 4 | US-09-299-549-4 |
| 22 | 147.5 | 5.2 | 492 | 2 | US-08-355-844-3 |
| 23 | 147.5 | 5.2 | 492 | 5 | PCT-US95-16126-3 |
| 24 | 147 | 5.2 | 488 | 2 | US-08-928-692-10 |
| 25 | 145 | 5.1 | 563 | 2 | US-09-031-392-2 |
| 26 | 145 | 5.1 | 563 | 4 | US-09-299-549-2 |
| 27 | 138 | 4.8 | 423 | 2 | US-08-494-907-14 |

| | | | | | |
|----|-------|-----|------|---|-------------------|
| 28 | 138 | 4.8 | 423 | 5 | PCT-US96-10986-14 |
| 29 | 135.5 | 4.8 | 383 | 2 | US-09-031-392-3 |
| 30 | 135.5 | 4.8 | 383 | 4 | US-09-299-549-3 |
| 31 | 133 | 4.7 | 322 | 4 | US-08-964-127-6 |
| 32 | 131.5 | 4.6 | 457 | 2 | US-08-882-704A-6 |
| 33 | 124 | 4.4 | 286 | 4 | US-08-964-127-4 |
| 34 | 116 | 4.1 | 3033 | 1 | US-07-925-695-5 |
| 35 | 112 | 3.9 | 109 | 2 | US-08-647-397-4 |
| 36 | 112 | 3.9 | 455 | 1 | US-08-035-928-2 |
| 37 | 111.5 | 3.9 | 1820 | 3 | US-07-998-289B-8 |
| 38 | 111.5 | 3.9 | 2100 | 2 | US-08-808-793-23 |
| 39 | 111.5 | 3.9 | 2100 | 3 | US-08-772-512A-19 |
| 40 | 109.5 | 3.8 | 418 | 4 | US-09-030-267-5 |
| 41 | 108 | 3.8 | 1976 | 3 | US-09-024-020B-9 |
| 42 | 108 | 3.8 | 1978 | 3 | US-09-024-020B-3 |
| 43 | 108 | 3.8 | 1988 | 3 | US-09-024-020B-4 |
| 44 | 106 | 3.7 | 569 | 2 | US-08-750-723A-2 |
| 45 | 106 | 3.7 | 569 | 4 | US-09-191-275-2 |

ALIGNMENTS

RESULT 1
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-501-572-3

Query Match 15.1%; Score 429; DB 3; Length 555;
Best Local Similarity 29.4%; Pred. No. 3.6e-39;
Matches 126; Conservative 70; Mismatches 157; Indels 76; Gaps 14;
QY 119 SWQVALITSVYFVGMSSSTLWGNISDQYGRKTKLTKISLVLTLYYGILSAFAPVYSWTLV 178

MOLECULE TYPE: peptide
US-09-040-444-3

Query Match 15.1% Score 429; DB 3; Length 555;
Best Local Similarity 29.4%; Pred. No. 3.6e-39;
Matches 126; Conservative 70; Mismatches 157; Indels 76; Gaps 14;

119 SWQVALLTSVVVGMSSSTLWGNISDQYGRKTKLKSIVLWTLVYGYLSAFAPVYSWLV 178
146 SWMLDLFOSSVNVGFFTCGSMISGVIADRFGRKCLLTTLVINAAGVLMASIPYTWMLI 205
179 LR---GLV---GFGIGGVPOSVTLIAEFLPMKARAKCILLIEVFWAIGTVFEVLA--VF 230
206 FRLIQGLVSKAGWLIIGYI-----LITEVVGGRYRTVGIQVAVTVGL---LVLAGVAY 257
231 VMPSLGHRWLLILSAVPLLLFAVLCFWLPESARYDVLGNOEKATATLKRIATENGAPMP 290
258 ALPH--WRWLOFTVALPNFFLLYYWCIPESPRWLIISONKNAEAMRIIKHIAKNGKSLP 315
291 LCKLIISROEDRGK-----MRDLF--TPHFRWTTLLLMFIFNSAFSYYGLVLLTTELFQA 344
316 ASLQRLRELEETGKKNLPSFLDLVTPQIRKHTMILMYNFTSSVLYOGLIM---HMGLA 372
345 GDVCGISSRKKRAVEAKCSLACEYI SEEDYMDLLWTTLSEFFPGVLTWIIDRLGRKKTMA 404
373 GD-----NIYLDFFYSALVEFFAFAFMIIILIDRIGRRYPWA 408
405 LC-FVIFSFCSLLLF-----CVGRNVLTLLLFIAFAFISGGFOAAVYVTPEV 451
409 ASNWVAGAACLASVFI PGDLOWLKIISCLGRMGITM-----AYEIVCLVNAEL 457
452 YPTATRALGLGTCSGMARVAGALITPFIQVLMLESSVYLTAVYSGCCLLAALASCFLPIE 511
458 YPTFIRNLGVHICSSMCDIGGIITPFLVYRLTNWLELPLVMFVGLVAGGLALLLP-E 516
512 TKGGGLQES 520
517 TKGRALPET 525

RESULT 3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

119 LR---GLV---GFGIGGVPOSVTLIAEFLPMKARAKCILLIEVFWAIGTVFEVLA--VF 230
146 SWMLDLFOSSVNVGFFTCGSMISGVIADRFGRKCLLTTLVINAAGVLMASIPYTWMLI 205
179 LR---GLV---GFGIGGVPOSVTLIAEFLPMKARAKCILLIEVFWAIGTVFEVLA--VF 230
206 FRLIQGLVSKAGWLIIGYI-----LITEVVGGRYRTVGIQVAVTVGL---LVLAGVAY 257
231 VMPSLGHRWLLILSAVPLLLFAVLCFWLPESARYDVLGNOEKATATLKRIATENGAPMP 290
258 ALPH--WRWLOFTVALPNFFLLYYWCIPESPRWLIISONKNAEAMRIIKHIAKNGKSLP 315
291 LCKLIISROEDRGK-----MRDLF--TPHFRWTTLLLMFIFNSAFSYYGLVLLTTELFQA 344
316 ASLQRLRELEETGKKNLPSFLDLVTPQIRKHTMILMYNFTSSVLYOGLIM---HMGLA 372
345 GDVCGISSRKKRAVEAKCSLACEYI SEEDYMDLLWTTLSEFFPGVLTWIIDRLGRKKTMA 404
373 GD-----NIYLDFFYSALVEFFAFAFMIIILIDRIGRRYPWA 408
405 LC-FVIFSFCSLLLF-----CVGRNVLTLLLFIAFAFISGGFOAAVYVTPEV 451
409 ASNWVAGAACLASVFI PGDLOWLKIISCLGRMGITM-----AYEIVCLVNAEL 457
452 YPTATRALGLGTCSGMARVAGALITPFIQVLMLESSVYLTAVYSGCCLLAALASCFLPIE 511
458 YPTFIRNLGVHICSSMCDIGGIITPFLVYRLTNWLELPLVMFVGLVAGGLALLLP-E 516
512 TKGGGLQES 520
517 TKGRALPET 525

RESULT 2
US-09-040-444-3
; Sequence 3, Application US/09040444
; Patent No. 6063786
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (202)408-4000
;
; TELEFAX: (202)408-4400
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 556 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: peptide
;
US-08-501-572-1

```

[illegible]

```

1  RESULT          4
2  US-09-040-444-1
3  : Sequence 1, Application US/0904044-1
4  : Patent No. 6063766
5  : GENERAL INFORMATION:
6  : APPLICANT: Koepsell, Hermann
7  : APPLICANT: Gorboulev, Valentin
8  : TITLE OF INVENTION: Transport p
9  : TITLE OF INVENTION: Transport O
10 : TITLE OF INVENTION: DNA Sequenc
11 : NUMBER OF SEQUENCES: 6
12 : CORRESPONDENCE ADDRESS:
13 : ADDRESSEE: Finnegan,Henderson
14 : STREET: 1300 I Street, N.W.,
15 : CITY: Washington
16 : STATE: D.C.
17 : COUNTRY: USA
18 : ZIP: 20005-3315
19 : COMPUTER READABLE FORM:
20 : MEDIUM TYPE: Floppy disk
21 : COMPUTER: IBM PC compatible
22 : OPERATING SYSTEM: PC-DOS/MS-D

```

```

: SOFTWARE: PatentIn release #1.0, Version #1.3.0
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/09/040,444
:
: FILING DATE: March 19, 1998
:
: CLASSIFICATION:
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: O'Connor, Steven P
:
: REGISTRATION NUMBER: 41,225
:
: REFERENCE/DOCKET NUMBER: 2481.1453-01
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (202)408-4400
:
: TELEFAX: (202)408-4400
:
: INFORMATION FOR SEQ. ID NO.: 1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 556 amino acids
:
: TYPE: amino acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: peptide
:
: US-09-040-444-1

```

| Query Match | 15.08; | Score 426.5; | DB 3; | Length 556; |
|-----------------------|--------------|---|------------|-------------|
| Best Local Similarity | 29.08; | Pred. No. 6.8e-39; | | |
| Matches 126; | Conservative | | Indels 65; | Gaps 12; |
| QY | 119 | SHQVALLTVSVFVGMSSSTLWGNISDOYGRGTGLKISLVLTWLYYGIILSAFAVYVSMILV | 178 | |
| DB | 146 | ANKVDLFQSCVNLGFFGLSLVVGVIADRFGRKLCLLVTLTVTSVSGVLTAVAPDYTSMLL | 205 | |
| QY | 179 | LRGLVGF-GIGVQPOSVTLVYAEFLPMKARAKCILLIEVFNAIGTVVEVVLAVFVMSLGM | 237 | |
| DB | 206 | FRLLQGMVSKGSWVSGYTLITEFVSGYRRTAILYQMAFTVGLVGLAGVA-YAIPD--W | 262 | |
| QY | 238 | RWLLILSAVPLLFAVLFCWLPESARYDVLVSGNOEKAIATLKRATENGAMPGLKGLIS | 297 | |
| DB | 263 | RWLQIAVSLPTFLFLYYVFWPESPESRWLLSQKRTTRAVRIMEQIAQNGKVPADLKMLC | 322 | |
| QY | 298 | ROEDRGKMR-----DLF-TPHFRWTTLLLPWFISNFAFVYGLVLLTTELFQAGDVCGIS | 351 | |
| DB | 323 | LEEDASEKRSFADLFRTNLRKHTVILMYLWVSCAVLYQGLIM----- | 367 | |
| QY | 352 | SRKKAVEAKCSLACEVLI-DEEDYMDLLMTLTSEFPGVLTVLWLIIDRLGRKKTMLAC-FVIF | 410 | |
| DB | 368 | -HVGATGANL-----YLDFYSSLVEFFPAAFILVTDIDRIGRIYPIAASNLVGTG | 415 | |
| QY | 411 | SFCSLLFTI-----CVGNRVLTLLLTARAFISGGFOAAVYVTEVPYPTATRA | 458 | |
| DB | 416 | AACLLMIFIPHELHNLVNTLACLRMGATIVL-----EMVCLVNAELYPPTFIRN | 464 | |
| QY | 459 | LGLGTCSGMARVGA LITPFIAQVMLESSVYLTAVYSGCCLLAALASCFPIPIETKGGGLQ | 518 | |
| DB | 465 | LGMWVCSALCDJGGITFPFWFRLEWVWQALPLILFVGLGTAGAMTLLLP-ETKGVALP | 523 | |
| QY | 519 | ESSHREWGQEMVGR | 532 | |
| DB | 524 | ETTER---AENLGR | 534 | |

```

RESULT 5
US-08-501-572-2
; Sequence 2, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport P
; TITLE OF INVENTION: Transport P
; TITLE OF INVENTION: DNA Sequenc
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson

```

STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/501.572
FILING DATE: 2000-05-31
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Toohy, Kimberlin M
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-501-572-2

Query Match 15.0%; Score 425.5; DB 3; Length 553;
Best Local Similarity 30.2%; Pred. No. 8.7e-39;
Matches 127; Conservative 66; Mismatches 167; Indels 61; Gaps 11;

Qy 119 SVOALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLTLYGILSAFAPVYSWILV 178
Db 145 SWKLDLFQSLNAGFFGSLGVGYFADRGKRLCLLGTVLVNAVSGVLMFSPNYMSMLL 204
Qy 179 LRGLVGF-GIGVPOSVTLVYAEFLPMKARAKCILLIEVFMAIGTVFEVVLAVFVMPSLGW 237
Db 205 FRLLQGLVSKGNMAGYTLITEFVSGSRRTVAIMYQMAFTVGLVALTGLA-YALPH--W 261
Qy 238 RWLLILSAVPLLLFAVLCFPLPESARYDVLGNOEKAIATLKRIATENGAPMLGLKLIIS 297
Db 262 RWLQVAVSLPTFLFLYYWCVPSPRWLLSQKRNTAEIKIMDHIAOKNGKLPADLKMLS 321
Qy 298 RQED-----RGKRDLP-TPHFRWTTLLWFIMFSNAFSGYGLVLTTELFOAGDVCGIS 351
Db 322 LEEDVTEKLSFADLFTPRLRKRTEFLMYLWFTDSVLYQGLIL-----HMGATSG-- 373
Qy 352 SRKKAWEAKCSLACEYLSEEDYMDLLWTTLSSEFPGVLTWIIDRLGRKKTAL----- 405
Db 374 -----NLYDLFLYSALVEIPGAFIALITIDRVGRIPMAVSNLLAG 414
Qy 406 --CFVIFSFCSL-----LLFICVGRNVLTLLFIARAFISGGFQAAYVYTPVPTATRAL 459
Db 415 AACLVIFISPDHLWLNIIIMCVGRMGITIAI-----OMICLVNAELYPTFVRNL 463
Qy 460 GLGTCSGMARVAGALITPFIAQVLMLESSVYLTNAVYSCCLLAALASCFLEPIETKGGLOE 519
Db 464 RVWVCSSICDGGIITPFIVFRREVWQALPLILFAVLGLLAAGVTLLLP-ETKGDALPE 522
Qy 520 S 520
Db 523 T 523

RESULT 6
US-09-040-444-2
; Sequence 2, Application US/0904044
; Patent No. 6063766
; GENERAL INFORMATION:

APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
TITLE OF INVENTION: DNA sequences encoding it and their use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040.444
FILING DATE: March 18, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-040-444-2

Query Match 15.0%; Score 425.5; DB 3; Length 553;
Best Local Similarity 30.2%; Pred. No. 8.7e-39;
Matches 127; Conservative 66; Mismatches 167; Indels 61; Gaps 11;

Qy 119 SVOALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLTLYGILSAFAPVYSWILV 178
Db 145 SWKLDLFQSLNAGFFGSLGVGYFADRGKRLCLLGTVLVNAVSGVLMFSPNYMSMLL 204
Qy 179 LRGLVGF-GIGVPOSVTLVYAEFLPMKARAKCILLIEVFMAIGTVFEVVLAVFVMPSLGW 237
Db 205 FRLLQGLVSKGNMAGYTLITEFVSGSRRTVAIMYQMAFTVGLVALTGLA-YALPH--W 261
Qy 238 RWLLILSAVPLLLFAVLCFPLPESARYDVLGNOEKAIATLKRIATENGAPMLGLKLIIS 297
Db 262 RWLQVAVSLPTFLFLYYWCVPSPRWLLSQKRNTAEIKIMDHIAOKNGKLPADLKMLS 321
Qy 298 RQED-----RGKRDLP-TPHFRWTTLLWFIMFSNAFSGYGLVLTTELFOAGDVCGIS 351
Db 322 LEEDVTEKLSFADLFTPRLRKRTEFLMYLWFTDSVLYQGLIL-----HMGATSG-- 373
Qy 352 SRKKAWEAKCSLACEYLSEEDYMDLLWTTLSSEFPGVLTWIIDRLGRKKTAL----- 405
Db 374 -----NLYDLFLYSALVEIPGAFIALITIDRVGRIPMAVSNLLAG 414
Qy 406 --CFVIFSFCSL-----LLFICVGRNVLTLLFIARAFISGGFQAAYVYTPVPTATRAL 459
Db 415 AACLVIFISPDHLWLNIIIMCVGRMGITIAI-----OMICLVNAELYPTFVRNL 463
Qy 460 GLGTCSGMARVAGALITPFIAQVLMLESSVYLTNAVYSCCLLAALASCFLEPIETKGGLOE 519
Db 464 RVWVCSSICDGGIITPFIVFRREVWQALPLILFAVLGLLAAGVTLLLP-ETKGDALPE 522
Qy 520 S 520

| | | | | |
|--|---|---|-----------|--|
| | D | b | 523 T 523 | |
| <RESULT 7 | | | | |
| US-08-647-397-2 | | | | |
| ; Sequence 2, Application US/08647397 | | | | |
| ; Patent No. 5972702 | | | | |
| ; GENERAL INFORMATION: | | | | |
| ; APPLICANT: Beier, David R. | | | | |
| ; APPLICANT: Brady, Kevin P. | | | | |
| ; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER | | | | |
| ; NUMBER OF SEQUENCES: 6 | | | | |
| ; CORRESPONDENCE ADDRESS: | | | | |
| ; ADDRESSEE: Wolf, Greenfield & Sacks, P.C. | | | | |
| ; STREET: 600 Atlantic Avenue | | | | |
| ; CITY: Boston | | | | |
| ; STATE: MA | | | | |
| ; COUNTRY: USA | | | | |
| ; ZIP: 02210 | | | | |
| ; COMPUTER READABLE FORM: | | | | |
| ; MEDIUM TYPE: Floppy disk | | | | |
| ; COMPUTER: IBM PC compatible | | | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | | | |
| ; SOFTWARE: PatentIn Release #1.0, Version #1.25 | | | | |
| ; CURRENT APPLICATION DATA: | | | | |
| ; APPLICATION NUMBER: US/08/647,397 | | | | |
| ; FILING DATE: | | | | |
| ; CLASSIFICATION: 424 | | | | |
| ; ATTORNEY/AGENT INFORMATION: | | | | |
| ; NAME: Gates, Edward R. | | | | |
| ; REGISTRATION NUMBER: 31,616 | | | | |
| ; REFERENCE/DOCKET NUMBER: B0801/7048 | | | | |
| ; TELECOMMUNICATION INFORMATION: | | | | |
| ; TELEPHONE: 617-720-3500 | | | | |
| ; TELEFAX: 617-720-2441 | | | | |
| ; INFORMATION FOR SEQ ID NO: 2: | | | | |
| ; SEQUENCE CHARACTERISTICS: | | | | |
| ; LENGTH: 537 amino acids | | | | |
| ; TYPE: amino acid | | | | |
| ; TOPOLOGY: linear | | | | |
| ; MOLECULE TYPE: protein | | | | |
| US-08-647-397-2 | | | | |
| Query Match 13.9%; Score 395; | | | | |
| Best Local Similarity 23.6%; | | | | |
| Matches 130; Conservative 91; Mismatches 194; Indels 136; Gaps 20; | | | | |
| QY 72 DAVEAIG-FKFKFWKLVSVLGLAWADAME--MMILSLAPOLHCE-----WRIP- 118 | | | | |
| : : : : : : : : : : : : : : : | | | | |
| D b 5 EILDVSGMGPFOYLHVTLALPLGIANHNLLQIFATTPDHHCRRPPNPASLEPWWLPL 64 | | | | |
| : : : : : : : : : : : : : : | | | | |
| QY 119 -----SMOVA-----L 124 | | | | |
| : : : : : : : : : : : : : : : | | | | |
| D b 65 GPNCKPEKLFVFHLPNASILPDNTQGATEPCLCDGIWYNSTRDTIVTEWDLCVGSKNLKEM 124 | | | | |
| : : : : : : : : : : : : : : | | | | |
| QY 125 LTVSVFVMSSSTLGNIISDQYGRTGLKIISLVLTLYGYLSAFPAYYSIWILVRGLVG 184 | | | | |
| : : : : : : : : : : : : : : | | | | |
| D b 125 AQSVFMAGILVGGPVFGELSDRDFGRKKPILTWSYLLLAAAGSAAFPSPSLTYVMIFRLCG 184 | | | | |
| : : : : : : : : : : : : : : | | | | |
| QY 185 FGIGGVPOS-VTLYAEFLPMKARAKILLIEVFWAICTVEFWLAFFVMPSLG-----WR 238 | | | | |
| : : : : : : : : : : | | | | |
| D b 185 CSISGISLSITLNWEVPYSTRAISTTTGYCYTIQO-----FILPLAVAVPWQR 236 | | | | |
| : : : : : : : : : : : : : : | | | | |
| QY 239 WLLIISLAPVLLLFAVCFLWPESARYDVLSGNOKAIATLRKRIAT-----ENGAPMLPGK 293 | | | | |
| : : : : : : : : : : : : : : | | | | |
| D b 237 WLQLSVSAAFFIESLLSWVPESISRWLVLSGFKRALKTKLRVATFNKGKEEGEKLTVEE 296 | | | | |
| : : : : : : : : : : : : : : | | | | |
| QY 294 LIISRQED-----RGKMRLF-TPHFRWTWLWFNFWSNAFSYYGLVLIITTELFAQGD 346 | | | | |
| : : : : : : : : : : | | | | |
| D b 297 LKFNLQKDITSARKVKYGLSDLFRVSIILRVVTFCULSLAMFATGFAYYSIAMGEVEEP----- 351 | | | | |
| : : : : : : : : : : | | | | |

```
QY 133 MMSSTLWGNISDOYGRKTKLISVLTWLYGILSAFAPVYSWILVLRGLVGFIC-GVP 191
DB 139 ALITISFAGTAADIFGRKCLMGSNLMFVIGAILQVSAHTEFWMAGVRLIMFGVGIGSL 198
QY 192 QSVTLVAEFLPMKARAKCILLIEVEFWAIGTVFEVVL-----AVFVMPSLGWRWLLIISAVP 247
DB 199 IAPLFISELAPKIRGR-LTVINSLWLTGG--OLVAYGGAGLNVVNGRMLVGLSLIP 255
QY 248 LLL-FAVLFCWLPESARYDVLGNOGKAIATLKRATENGAPM---PLGKLLISROEDRG 303
DB 256 TAVQFTCLCF-LPDTPRYVMKGLARATEVLRKSYTDTSEEIIRKVFELVTLNQSIPG 314
QY 304 KMRDLTPHERWTLTLLWFIFWESNAPSYGLVLLTTELFQAGDVCGISSRKKAVEAKCSL 363
DB 315 KN-----VPEKVNWTI-----KELHTVPSNLRALIIGCOL 344
QY 364 -ACEYLSEEDYMDLLWTLTSEFPG-----VLVTLWIIDRL 397
DB 345 QAIQOFTG-----WNSLMYSGTIFETVGRKNSAVSIIVSGTNFIETLVAFSSIDKI 397
QY 398 GRK-----KTMAL--CFVIFSPC-----SLLLFICVGRN-----VLTLLFTIARAF 436
DB 398 GRRTILLIGPGMTALVWCSTAFHGLKFDGCAVAVVSSGFWIIVFIIVFAAF 457
QY 437 ISGFOAAVYVTPVYPTATRALGLGTCGSMARVAGALITPFAQVNMLESVVYL-TLAVYS 495
DB 458 YALGIGTVPMQOSELFPQNVRGIGTSYATATNAGSLVIASITFLTMLQNTITPAGTFAPFA 517
QY 496 GCCLLAALASCFLPIETKGGGLQE 519
DB 518 GLCLSTIFCYFCYPELSGLEEE 541
```

RESULT 9

```
US-08-964-127-2
; Sequence 2, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
; REFERENCE/DOCKET NUMBER: 07334/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 520 amino acids
; TYPE: amino acid
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-964-127-2
```

```
Query Match 7.2%; Score 203.5; DB 4; Length 520;
Best Local Similarity 25.6%; Pred. No. 5.7e-14;
Matches 116; Conservative 68; Mismatches 174; Indels 95; Gaps 24;
```

```
QY 104 LSILAPQLHCEWRLP---SQWVALLTSVYFVGMSSSTLWGNISDOYGRKTKLISVLTW 160
DB 79 LPVLTNAIGQWDLVCDLQWQILEQILFILGFASGYFLGYPADRFGR--GI---VLIT 134
QY 161 LVY-----GILSAFAPVYSWILVLRGLVGFICGVPQSVTLV-AEFLPMKARAKCILLIEV 215
DB 135 LGLVPCGVGGAAAGSSTGVMALELLGFLLAGVDLGVYLMRLELCDPTQRLKVALAGEL 194
QY 216 FWAICTVFEVLAVFVMPSLGWRWLLIISAVPLLIIFAVLCFWLP-----ESARYDVLG 271
DB 195 ---VGVGHFLEGLALVSKDMRFLOQMITAPCILF--LFYGMWPGFLFESARWLVKROI 249
QY 272 EKAIATLKRATENGAPMPLGKLLISROEDRGKMRDL-----FTPHFRWTLTLL---L 320
DB 250 EEAQSVLRILABRN---RPHGQML--GEEAQALQDLENTCPLPATSSFSFASLLNRYNI 304
QY 321 W-----FIMFSNAFYSYGLVLLTTELFQAGD-----VCGISSRKKAVEAKCSLACEY 371
DB 305 WKNLLILGFTN-FIAHAIRHCYQPVGGGSPSDFYLCSLASGTA-----ALACVFL--- 355
QY 372 DYMDLLWTLTSEFPGVLTWLIIDRLGRK..MALCFVIFSPCSLLLLFICVGRNVTLLIF 431
DB 356 -----GVTV-----DRFRRGILLISMTLTGIASLVL-----LGLWDY 388
QY 432 IARAFIS-----GGF--QAAYVVT-----PEVYPTATRALGLGTCGSMARVAGALITP 480
DB 389 LINEAALTFSVLGLFSSQAAAILSTLLAAEVIPTVRGKGLGLMALGALGGLSGP-AQR 447
QY 481 VMLESVVLTAVYSGCCLLAALASCFLPIETK 513
DB 448 LHMHGAFLOHVVLAAACALLCILSIMLLP-ETK 479
```

RESULT 10

```
US-08-928-692-12
; Sequence 12, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
```

```

;
;   REGISTRATION NUMBER: 33,728
;   REFERENCE/DOCKET NUMBER: 4944,200-US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-867-0123
;   TELEFAX: 212-878-9655
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 524 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: No. 5958727e
;
US-08-928-692-12

Query Match      7.2%; Score 203.5; DB 2; Length 524;
Best Local Similarity 22.0%; Pred. No. 5.7e-14;
Matches 119; Conservative 81; Mismatches 197; Indels 143; Gaps 22;

QY 48 VELDDGAAPKEFANPTDD-----TFMVEDAVEAIGFGKQKLSVLTGLAW 94
DB 48 VPLDDRKAANNVINSTDELPTTISYNNPKPTPAEEETVAAAALITMLWSLV----- 101
QY 95 MADAMEMMILSIAPLQHCWRPLSNQVALLTSVVFVGMSSSTLMGNISDOYGRKTGLK 154
DB 102 -----SSFAVGGMGTASFFGGWGLDGLTGRKAML 129
QY 155 ISVLWTLYYGILSAFA--PVYSMWILVRLGLVGGVGPVOSTVLY-AEFLPMKARAKCI 210
DB 130 VANILSLVGLALLMGFKSLGPHSHILITAGRSISGLYGLISGLVPMYIGEIAPTALRG-- 186
QY 211 LLIEVFMAIGTFVEVVLAVFMPS-----LG----WRWLLILSAVPLLLFAVLFCFWL 258
DB 187 -----ALGTFHQLAIVTGILISIIITGLEFILCNVDLHILLGLSGVRAILQSLLIFFC 239
QY 259 PESARYDVLSONQE-KAIAATLKRIATENGAPMPLGLKLIISROEDRGKMR-----DLFT-PH 312
DB 240 PESPRYLKILDEEVKAKSLRKLRGVDDVTKDINEMKEREASQKVSIIQLFTNSS 299
QY 313 FRWTTLLWFTWFSNAPS-YVGLVLLTTELFQAGDVGCGISSRKKKAVEAKCSLACEYLSEE 371
DB 300 YROPILVALMLHVAQOFGSINGIFYSTSIFQ---TAGIS---KPVYATIGVGA----- 347
QY 372 DYMDLLTTLTSEFPGLVLTWIDRLGRKKTMAFCVIFSFCSLLLFICVGR----- 423
DB 348 ---VNMVTFVSVF-----LVEKAGRRSLFLIGMSGMFVCA--IFMSGVLVLLNKFWS 395
QY 424 ----NVLTLLLFIARAFTSGG----FOAAVYVT--PEVYPTATRALGLGTCGSMARVYAL 473
DB 396 MSYVSMIAIFLVSFSEFGPGPIPMFWVAEFFSGQPRPAALATAAFSNWTCNFIV---AL 452
QY 474 ITPPIAQWMLSESVYLVAVYSGCCLLAALASCFLPIETKGGGLOE-----SSHR 523
DB 453 CFQYIADFCGPYVFFL-----FAGVLLAFTLTFFKVPETKGSFEETAFAEFQKKSAAHR 508

RESULT 11
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
```

```

;
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows95
;   SOFTWARE: FastSeq for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/031,392
;   FILING DATE: 26-FEB-1998
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Meiklejohn, Ph.D., Anita L.
;   REGISTRATION NUMBER: 35,283
;   REFERENCE/DOCKET NUMBER: 07334/072001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
;   TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 493 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-09-031-392-10

Query Match      6.3%; Score 178; DB 2; Length 493;
Best Local Similarity 21.7%; Pred. No. 3.7e-11;
Matches 110; Conservative 92; Mismatches 196; Indels 108; Gaps 23;

QY 67 TFMVEDAVEAIGFGKQKLSVLTGLAWMADAMEMMILSIAPLQHCWRPLS----- 119
DB 13 TLVLAVLIAALGSGFOYGNLGV-----NAPQKVEAFYETWLGXGXSPVTLTL 65
QY 120 MQVALLTSVVFVGMSSSTLMGNISDOYGRKTGLKLSVLWTLYYGILSAFAPV---YSWI 176
DB 66 W--SLSVISFVAGMGTGSEFLVCGXIGNRLGRKXAMLVNNVLATAGGLJMLGLAKXAXSFEML 123
QY 177 LVLRGLVGGVGGVGPVOSTVLY-AEFLPMKARAKCILLIEVFMAIGTFVE--VVLAVFVMP 233
DB 124 ILGRFTIIGLYCGISSGVPMYVGEISPTALRG-----ALGTNLQGLIVIGILIAQ 173
QY 234 SLG-----WRWLLILSAVPLLLFAVLFCWLPESARYDVLSONQE-KAIAATLKRIA 282
DB 174 VLGLDSLGNESLWPLLLGLTGVPAQLLLLPFCPESPRYLLINKNEAKAKKALQRLR 233
QY 283 TENGAMPPLGKLIISROEDRG-----KMRDLF-TPHRWTTLLWFWFS-----NA 328
DB 234 GTADVSOEVAEM--KDESXMXSEKXSVLELFRSRXYRQPVIIAIVLQSLQSLSCINA 290
QY 329 FSYVGLVLTTELFQAGDVGCGISSRKKKAVEAKCSLACEYLSEEDYMDLLTTLTSEFPGLV 388
DB 291 VFYV---STSIKAGV-----QPVYATIGAG-----VNMVTFVSVF----- 327
QY 389 VTWIDRLGRKKTMAFCVIFSFCSLLLFICVGR-----NVLTLLLFIARAFTS 438
DB 328 ----VVERAGRRTLLHLLGLGGMAGCAVLMTIALALLDQVPMMSVYSIVAIFFGVAFVFG 383
QY 439 GG-----FOAAVYVTPEVYPTATRALGLGTCGSMARVYALITPTFAQWMLSESVYLVAVY 494
DB 384 PGPIPMFIVAELEFSQGRPAIAVAGFSNNWTSNFIVG-LLFQYIAE-LLGPGYVFIQVAFV 440
QY 495 SGCCLLAALASCFLPI-ETKGGGLOE 519
DB 441 ---LLLLFFITFLKVPETKGRTFDE 463

RESULT 12
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
```

;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
;; TITLE OF INVENTION: GLUTEX AND USES THEREOF
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 225 Franklin Street
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows95
;; SOFTWARE: FASTSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/299,549
;; FILING DATE: 26-APR-1999
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/031,392
;; FILING DATE: 26-FEB-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meiklejohn, Ph.D., Anita L.
;; REGISTRATION NUMBER: 35,283
;; REFERENCE/DOCKET NUMBER: 07334/072002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 493 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-299-549-10

Query Match 6.3%; Score 178; DB 4; Length 493;
Best Local Similarity 21.7%; Pred. No. 3.7e-11;
Matches 110; Conservative 92; Mismatches 196; Indels 108; Gaps 23;
QY 67 TMVEDAVEAIGFGKQKLSVLTGLAMMADAMEMMILSILAPQLHCEWRPLFS-----119
Db 13 TLVLAVLIAALGSGFYGYNLGVI-----NAPQKVEAFYETWLGXGXPVPTLTLL 65
QY 120 WQVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLWLTLYYGILSAFAPV---YSWI 176
Db 66 W--SLVSIFAVGGMIGSFLVXIGNRLGRKXAMLVNNVLAITAGGLMLGLAKXAXSFEML 123
QY 177 LVLRLGVGFGIGVQPSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFE--VVLAVFVMP 233
Db 124 ILGRFIIGLYCGLSGGVPMYVGEISPTALRG-----ALGTLNQLGIVIGILLAQ 173
QY 234 SLG-----NRWLLILSAVPLLLFAVLFCWLPESARYDVLGNGOE-KAIAITLKRIA 282
Db 174 VLGLDSLLGNESLWPLLLGLTGPALLQLLLPFCPSRYLLINKNEARAKKALQRLR 233
QY 283 TENGAPMPGLKLIISROEDRG-----KMRDLF-TPHFRWTTLLWFIWFS-----NA 328
Db 234 GTADVSQVEAEM--KDSRXMXSEKXSVLELFRSRYRQPVIIAIVLSQLSGINA 290
QY 329 FSYGLVLLTTLFQAGDVCGISSRKKAVEAKCSLACEYLSIEDYMDLLWTTLSFPGVL 388
Db 291 VFY-----STIFEKAGV-----GQPVYATIGAG-----VNVTVFTVVSVF---- 327
QY 389 VTLWIIDRLGRKTKHALCFVIFSCSLLLFICVGR-----NVLTLILFIARAFIS 438
Db 328 ----VVERAGRRTLHLGLGGMAGCAVMTIALALDDQVPMMSYVSIVAIFGVFAFFVG 383
QY 439 GG-----FOAAVYVTEVYPTATRALGLGTCGSMARGALITPTFAQVMESSVYLTAVY 494
Db 384 PGPIPWFIIVAEILFSQGPRAATAVAGFSNWTNSNFVIG-LLFOYIAE-LLGPVVFIVFAV- 440

QY 495 SGCCLLAALASCFLPI-ETKGGLOE 519
Db 441 ---LLLLFFIITFLKVPETKGRTFDE 463
RESULT 13
US-09-031-392-5
;; Sequence 5, Application US/09031392
;; Patent No. 5942398
;; GENERAL INFORMATION:
;; APPLICANT: Tartaglia, Louis A.
;; APPLICANT: Weng, Xun
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
;; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 225 Franklin Street
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows95
;; SOFTWARE: FASTSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/031,392
;; FILING DATE: 26-FEB-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meiklejohn, Ph.D., Anita L.
;; REGISTRATION NUMBER: 35,283
;; REFERENCE/DOCKET NUMBER: 07334/072001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 494 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-031-392-5

Query Match 6.0%; Score 171; DB 2; Length 494;
Best Local Similarity 23.7%; Pred. No. 2.3e-10;
Matches 123; Conservative 82; Mismatches 205; Indels 110; Gaps 27;
QY 73 AVEAIGFKQKLSVLTGLAMMADAMEMMILSILAPQLHCEWRPLFSWOVALLTS-----127
Db 13 ASIAITGFSFGYGN--TG---VINAPEAIKDFLNTYLEERSETPPSSV-LTSLWSLS 66
QY 128 --VVFVGMSSSTLWGNISDOYGRKTGLKISVLWLTLYYGILSAFAPVYSMILVL---RGL 182
Db 67 VAIFSVMGIGSFGVLFVNRGRNSMLIVNLLAITAGGLMGFCCKIAESVEMLILGLRI 126
QY 183 VQFGIGGVQPSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFE--VVLAVFVMPSLG---236
Db 127 IGLFCGLCTGTFVPMYIGEISPTALRG-----AFGLNQLGIVIGILVAQIFGLKV 176
QY 237 -----NRWLLILSAVPLLLFAVLFCWLPESARYDVLGNGOEKAIAITLKRI-APENGA 287
Db 177 ILGTEDLWPLLLGFPILFAIIQCAALPFCPSRFLINKKEEKAKEILQRLWGTEDVA 236
QY 288 ----PMPGLKLIISROEDRGKMRDLF-TPHFRWTTLLWFIWFS-----NAFSYGLVL 336
Db 237 QDIQEMKDESMMS-QEKQVTVLELFRAPNRPQIISIMLQLSQOLSGINAVFY----291
QY 337 LTTELFQAGDVCGISSRKKAVEAKCSLACEYLSIEDYMDLLWTTLSFPGVLVTLWIIDR 396

Db 292 -STGIFKDGAV-----QEPVYATIGAG-----VVNTIFTVSVF-----LVER 328
QY 397 LGRKKT-----MALCFVIFSCSLL-----FICVGRNVLTLLLLFIARAFISGG- 440
Db 329 AGRTLHLGLGMAFCSILMTI-SLLKDNYSWMSFICIG-----AILVFVAFEIGPGP 383
QY 441 ---FQAAVYVTPVPTATRALGLGTCGSMARVAGALITPFIQVLMLESSVYLILAVYSGC 497
Db 384 IPWFIVAEELFGQGRPAAMAVAGCSNWTSNFLVGLLFPSS--ATFYLGAYVFIVTFV- 438
QY 498 CLLAALASCLPIETKGGGLOESSHREWGOMVG-RGMHG 536
Db 439 LVIFWVTFKVPETGRTEETITRAFEQVQTGRGEGK 478

RESULT 14
US-09-299-549-5
; Sequence 5, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549
; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-299-549-5

Query Match 6.08; Score 171; DB 4; Length 494;
Best Local Similarity 23.7%; Pred. No. 2.3e-10;
Matches 123; Conservative 82; Mismatches 205; Indels 110; Gaps 27;

QY 73 AVEAIGFGKFKWKLVSITGLAWMADAMEMMILSILAPOLHCEWRLPSSQVALLTS----- 127
Db 13 AIIATIGSFQGYN--TG---VINAPAIKDFNLTLEERSETPPSSV-LLTSLWLS 66
QY 128 --VVFVGMSSLTWGNISDQYGRKTKIGLSVLWTLFYXGILSAFAPVYSWLVL---RGL 182
Db 67 VAIFSVGGMIGSFVGLFVNRFRGRNSMLIVNLLAIAGGLMGFKIAESVEMILIGRLI 126

QY 183 VGFGIGGVPOSVTLY-AEFLPMKARAKCILLIEVFWAIGTVE--VVLAVFVMPSLG--- 236
Db 127 IGLFCGLCTGFVPMYIGEISPTALRG-----AFGLNQLGIVIGILVAQIFGLKV 176
QY 237 -----WRWLLILSAVPLLFAVLCFWLPESARYDVL-SGNOEKAIATLKRI-ATENGA 287
Db 177 ILGTEDLWPLLGLFTILPAIIQCAALPFCPSPRFLLINRKEBEKAKEILQRLWGTEDVA 236
QY 288 ----PMLPGLKLIISROEDRGKMRDLF-TPHFRWTTLLLLWFIWES-----NAFSYVGLVL 336
Db 237 QDIQEMKDESMRS-QEKQVTVLELFRAPNYROPILISIMLQJLQSOQLSGINAVYY---- 291
QY 337 LTTLFQAGDVCGIGISSRKKAACEYLSSEEDYMDLLTTLSEFFPGVVLVTLIIDR 396
Db 292 -STGIFKDGAV-----QEPVYATIGAG-----VVNTIFTVSVF-----LVER 328
QY 397 LGRKKT-----MALCFVIFSCSLL-----FICVGRNVLTLLLLFIARAFISGG- 440
Db 329 AGRTLHLGLGMAFCSILMTI-SLLKDNYSWMSFICIG-----AILVFVAFEIGPGP 383
QY 441 ---FQAAVYVTPVPTATRALGLGTCGSMARVAGALITPFIQVLMLESSVYLILAVYSGC 497
Db 384 IPWFIVAEELFGQGRPAAMAVAGCSNWTSNFLVGLLFPSS--ATFYLGAYVFIVTFV- 438
QY 498 CLLAALASCLPIETKGGGLOESSHREWGOMVG-RGMHG 536
Db 439 LVIFWVTFKVPETGRTEETITRAFEQVQTGRGEGK 478

RESULT 15
US-08-928-692-11
; Sequence 11, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727 No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 488 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
US-08-928-692-11

```
Query Match      5.5%; Score 157; DB 2; Length 488;
Best Local Similarity 24.9%; Pred. No. 8.2e-09;
Matches 104; Conservative 66; Mismatches 198; Indels 50; Gaps 18;

QY 121 QVALLTSVVFVCGMSSSTLWGNISDQGRKTKLISVLWTLYYGILSAFAPVYWSILVLR 180
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 86 OYGAITSIFSIGLFGSYAGWANRYGRKYVSMGASAMCVSSLLFFFSNSYLQLLFGR 145
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 181 GLVFGIGGVQSVTLYA-EFLPMKARAKCILLIEVFWAIGTVFEVVLAVFVMPSLGHRW 239
    ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 146 FLVGHSCGTAIVITPLFNEIAPVEWCGAMGSMNQVSINLGGILLTQTILALKYADSYNWRW 205
    || |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 240 LILSAVPLLLFAVLCFWL--PESARYDVLGNQEKAIATLKRATENGAPMPLGKLIIS 297
    || |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 206 LLFSGSV--IAVANILAWLKYDESPRWLVSHGFVSEATLFLK--RPGTYQAKQEIQD 261
    | |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 298 ROEDRGKMRDLFTPHFRWTTLLLPWFMSNAPSYGLVLLTTELFQAGDVCGISS-RKKA 356
    | |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 262 WQRSHGNRPESSEETHSGPTLWQYVTDPSYKKPRTVILA--ILSCQCFCGINSIIFYG 319
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 357 VEAKCSLACEYLSEEDY-MDLLWTTLSBPPGVLVTL---WIIDRLGRKKTALCFVIFSF 412
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 320 VKVICKILPKYSIQNFASIL-----NVVTLAASAIIDHVGRRPLLLASTTVMTA 371
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 413 CSLLLFICVGRNVLTLL-----LFIARAFISGGFOAAVYVTPV-YP-TATRALGIGT-C 464
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 372 MSLLISVGLTLSVSFLVTATFVYIA-AFAIGLGPFPFLIIGELSYPODAATAQSGFTVC 430
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 465 SGMAR--VCALITPPIAQVMLESSVYLTAVYSGCCLLAALASCFLPI-----ETKG 514
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 431 NMLATFIVGYLPP--IGHGLMGGYVF-----AIFAAIAAMPATYVYKRVPETKG 477
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Search completed: March 13, 2002, 12:43:20
Job time: 186 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 25.88 Seconds
(without alignments)
1612.970 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 2846
Sequence: 1 MEEDLQLRLPVVKFRRTG.....MVGRGMHGAGVTRNSGSQGE 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 1280 | 45.0 | 529 | T23190 | hypothetical prote |
| 2 | 735.5 | 25.8 | 300 | S15786 | glucose transport |
| 3 | 734 | 25.8 | 144 | T50622 | hypothetical prote |
| 4 | 644.5 | 22.6 | 454 | T75580 | probable sugar tra |
| 5 | 509.5 | 17.9 | 422 | G72234 | hypothetical prote |
| 6 | 483 | 17.0 | 400 | G69757 | transporter homolo |
| 7 | 466.5 | 16.4 | 455 | B83213 | probable MFS trans |
| 8 | 436 | 15.3 | 593 | JC4884 | organic cation tra |
| 9 | 426.5 | 15.0 | 556 | S50862 | organic cation tra |
| 10 | 409 | 14.0 | 742 | A43344 | synaptic vesicle p |
| 11 | 403 | 14.2 | 742 | S27263 | synaptic vesicle p |
| 12 | 402.5 | 14.1 | 683 | S34961 | transmembrane tran |
| 13 | 395 | 13.9 | 724 | S50531 | probable MFS trans |
| 14 | 381.5 | 13.4 | 452 | A85852 | probable MFS trans |
| 15 | 366 | 12.9 | 446 | B83033 | probable MFS trans |
| 16 | 357.5 | 12.6 | 557 | JW0089 | organic cation tra |
| 17 | 355 | 12.5 | 448 | G83616 | 4-hydroxybenzoate |
| 18 | 335.5 | 11.8 | 557 | JE0346 | high-affinity earn |
| 19 | 334.5 | 11.8 | 576 | T22509 | hypothetical prote |
| 20 | 333.5 | 11.7 | 423 | S74046 | probable sugar tra |
| 21 | 333.5 | 11.7 | 448 | H83335 | probable MFS trans |
| 22 | 322 | 11.3 | 451 | S66008 | transport protein |
| 23 | 321.5 | 11.3 | 444 | E83033 | probable MFS trans |
| 24 | 311 | 10.9 | 457 | E70070 | metabolite transpo |
| 25 | 307 | 10.8 | 423 | A83383 | probable MFS trans |
| 26 | 297 | 10.4 | 401 | E69501 | sugar transporter |
| 27 | 296 | 10.4 | 450 | C83394 | probable MFS trans |
| 28 | 295.5 | 10.4 | 409 | T47026 | hypothetical prote |
| 29 | 291 | 10.2 | 469 | G65058 | hypothetical prote |

30 287.5 10.1 442 2 A83122 probable MFS trans
31 286.5 10.1 443 2 E64725 yaaU protein - Esc
32 285.5 10.0 443 2 H85485 probable transport
33 280.5 9.9 407 2 C64167 hypothetical prote
34 279.5 9.8 418 2 A64763 probable transport
35 279.5 9.8 418 2 D85530 probable transport
36 275 9.7 557 2 T38125 myo-inositol trans
37 267 9.4 464 2 F69587 L-arabinose transp
38 266 9.3 539 2 C96758 probablle protein
39 265.5 9.3 473 2 G69789 sugar transporter
40 265.5 9.3 794 2 T27870 hypothetical prote
41 265 9.3 511 2 A84537 hypothetical prote
42 264.5 9.3 547 2 A48442 membrane sugar tra
43 263 9.2 522 2 E86246 glucose transport
44 262 9.2 522 2 S12042 glucose transport
45 260.5 9.2 480 2 T23608 hypothetical prote

ALIGNMENTS

RESULT 1

T23190

hypothetical protein ZK637.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000

C:Accession: T23190

R:Craxton, M.

submitted to the EMBL Data Library, April 1993

A:Reference number: Z19704

A:Accession: T23190

A:Status: preliminary; translated from GB//MBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-529 <WIL>

A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1

A:Experimental source: clone K01F9

C:Genetics:

A:Gene: CESP:ZK637.1

A:Map position: 3

A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3

C:Superfamily: Caenorhabditis elegans glucose transport protein

| | | | | |
|-----------------------|-------|--|-------|------------------------------------|
| Query Match | 45.0% | Score 1280; | DB 2; | Length 529; |
| Best Local Similarity | 52.8% | Pred. No. 1.3e-97; | | |
| Matches | 258; | Conservative | 74; | Mismatches 125; Indels 32; Gaps 7; |
| Qy | 42 | HVG---LEAVELDDGAAPKPEANPTDD-----TFMVEDAVEAIGFGFOWKLSVLT | 90 | |
| Db | 34 | HVGDDFAVRYNLDSDT-----ELGPTDQRSPDSEKFTTVDVEAVGALGFGRFQLKLSLT | 89 | |
| Qy | 91 | GLANMADAMEMMILSLAPOLHCEWRPLPSWQVALLTSVYFVGMSSSTLMGNISDQYGRK | 150 | |
| Db | 90 | GMANMADAMEMMILSLIPALACENGISVQOALVTTCTVFGMMLSTFWGKICDRFGR | 149 | |
| Qy | 151 | TGLKISVLTWLYGILSAFAPYVSWILVRLGVGIGVQSVTLYAEFLPMKAKCI | 210 | |
| Db | 150 | KGLFTSTLVACIMGVISMSHPHYVLLFFRGLTGFGIGVQSVTLYAEFLPTQRAKCV | 209 | |
| Qy | 211 | LLIEVFVAIGTVFVFLAVFVMPSLGWRWLLILSLAPVLLFLFVLCFWLPESARDVLSGN | 270 | |
| Db | 210 | VLIESFAIGAVFEALLAYFVNESFGWRALMFLSSLPLGIFAVASFVLPESARFDMASGH | 269 | |
| Qy | 271 | QEKAIATUKRIATENGAMPMLGKLIISRQ----EDGKKMRDLTFPHFRWTTLLLLFWFSN | 327 | |
| Db | 270 | PERALETLQAAARNRVOLPTGRVLSSTKAGSESGDIANLLSPDLRKTITLLMCWIAIT | 329 | |
| Qy | 328 | AFSYGYGLVLLTTELOAGDVC--GISSRKKAVEAKCSLACEVLSYEDYMDLLTWTLSFFP | 385 | |
| Db | 330 | AFSYGMVLTFTVLFQSHDECHGLFSNGTQME-----VCOPLFRSDFTDLLSTLAEFP | 384 | |
| Qy | 386 | GVLVTLTIIDRLGRKKTMLACFVIFSCSLLLFTICVGRNVLTLTLFIARAFISGGFQAY | 445 | |

Wed Mar 13 13:08:14 2002

Db 385 GLIITVLIIEWFGKRTMALEYAVFAIFTLFLYFCLDRFTVTVLIFVARAFISGAFOQAY 444
 QY 446 VYTPVYPTATRALGLGTCGSMARVAGALITPFIAQVMLESSVYLTAVYSGCCLLAALAS 505
 Db 445 VYTPVYPTATRALGLGTCGSMARVAGALITPFIAQVMLESSVYLTAVYSGCCLLAALAS 497
 QY 506 CFLPIETKG 514
 Db 498 LSLPIETKG 506

RESULT 2
 S15786
 glucose transport protein homolog - Caenorhabditis elegans (fragment)
 C:Species: Caenorhabditis elegans
 C:Date: 06-Jan-1995 #sequence_revision 05-May-2000 #text_change 05-May-2000
 C:Accession: S15786
 R:Craxton, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; Hall
 Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.
 submitted to the EMBL Data Library, May 1991
 A:Reference number: S15786
 A:Molecule type: DNA
 A:Residues: 1-300 <CRA>
 A:Cross-references: EMBL:Z11115
 C:Genetics:
 C:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3
 C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 25.8%; Score 735.5; DB 2; Length 300;
 Best Local Similarity 55.1%; Pred. No. 4.6e-53;
 Matches 147; Conservative 38; Mismatches 67; Indels 15; Gaps 3;
 QY 42 HVG-----LEAVELDDGAAPKFEANPTDD-----TFWEDAVEAIGFGKQWKLSVLT 90
 Db 34 HVGDDFAVRYSNLDDRT----ELGEPTDQRSPDSEKFTTVDVAEALGFGFQLKSILT 89
 QY 91 GLAWMADAMEMTILSILAPOLHCEWRPLSPQVALLTSVVFVGMSSSTLWGNISDQYGRK 150
 Db 90 GMAMMADAMEMULLSILPALACENWISSVQALVTTVCFSMMLSSSTFWGKICDRFGR 149
 QY 151 TGLKISVLTLYGILSAPVYSWILVRLGLVFGIGVQVPSVLYAEFLPMKARAKI 210
 Db 150 KGLTFTFLVACIMGVISGMSPHFYVLLFRGLTGTGIGVQVPSVLYAEFLPTAQRACV 209
 QY 211 LLIEVFWAICTVEVVLAVFVMPISGLRWMLLISAVPLLLFVLCFVLPESARYDVLGN 270
 Db 210 VLIESFWAICAVEALLAVFVWESFGWRALMFLSSLPGLIFAVASFWLPESARFDMASGH 269
 QY 271 QEKAIATLKRIATENGAPMLGLIIS 297
 Db 270 PERALETQAARNRVQLPTGRLVSS 296

RESULT 3
 T50622
 hypothetical protein DKFZp761H039.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
 C:Accession: T50622
 R:Blum, H.; Bauersachs, S.; Wewes, H.W.; Weil, B.; Wiemann, S.
 submitted to the Protein Sequence Database, June 2000
 A:Reference number: Z25145
 A:Accession: T50622
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-144 <AAA>
 A:Cross-references: EMBL:AL359592
 A:Experimental source: adult amygdala; clone DKFZp761H039
 C:Genetics:
 A:Note: DKFZp761H039.1

Query Match 25.8%; Score 734; DB 2; Length 144;
 Best Local Similarity 99.3%; Pred. No. 2.7e-53;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 405 LCFVIFSCSLLLFICVGRNVLTLLLFIAFAFISGGFQAAYVYTPVYPTATRALGLGTC 464
 Db 1 LCFVIFSCSLLLFICVGRNVLTLLLFIAFAFISGGFQAAYVYTPVYPTATRALGLGTC 60
 QY 465 SGMARVAGALITPFIAQVMLESSVYLTAVYSGCCLLAALASCFPIETKGGGLQESSHRE 524
 Db 61 SGMARVAGALITPFIAQVMLESSVYLTAVYSGCCLLAALASCFPIETKGGGLQESSHRE 120
 QY 525 WQEWVGRGMHAGVTRNSGSQE 548
 Db 121 WQEWVGRGMHAGVTRNSGSQE 144

RESULT 4
 F75580
 probable sugar transporter - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: F75580
 R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
 S.; Shen, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: F75580
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-454 <WHI>
 A:Cross-references: GB:AE001863; GB:AE001825; NID:G6460670; PIDN:AAF12486.1; PID:964
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DRA0271
 A:Map position: 2
 C:Superfamily: yaaU protein

Query Match 22.6%; Score 644.5; DB 2; Length 454;
 Best Local Similarity 30.6%; Pred. No. 2.3e-45;
 Matches 142; Conservative 95; Mismatches 190; Indels 37; Gaps 5;
 QY 70 VEDAVEAIGFGKFWKLSVLTGLAWMADAMEMMILSILAPOLHCEWRPLSPQVAV--LLTS 127
 Db 15 VDRVLDLGLGRFQWKLKILCGLTWAADAMEVLLMGFALPGTSAAFELPKGSPAATMLLT 74
 QY 128 VVFCVGMSSSTLWGNISDQYGRKTKGLKISVLTLYGILSAPVYSWILVRLGLVFGFI 187
 Db 75 ATFAGLMFGAMFWGYLADRVGRSVFTTVALGVVFGLAGALAPTLTLLVAVRFLTGAI 134
 QY 188 GG-VPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVFLAVFY----MPSLGRWLLI 242
 Db 135 GGTLPVDYSMAAEFVPTAWRGFLVYLESFVAAGTVVAAALAWWYSTAFAPAEGWRLLIG 194
 QY 243 LSAPVLLFVLCFVLPESARYDVLGSGNOKAIAITLKRIATENGAPMLGLIISROEDR 302
 Db 195 LAALPGLVGLIARIIGTPDSPRSLLARGEQAARAAQVAAQAAQGTLPAAPLAHPEQPPR 254
 QY 303 GKMRDLFTPHFRMTTLLWFIWFSNAFSYGLV-----LLTTELFQAGDVCGISSRKAV 357
 Db 255 VSPAQLFRGVARRPTLLMVTWFGLSLGYGIFSWLPFLRAQGLDLGAV----- 304
 QY 358 EAKCSLACEYLSEEDYMDLWTLTLEFFQVLTWLIIDRLGRKKTMAICFVIFSCSLLL 417
 Db 305 -----YRSTLLALAQVPGYLLAAVLEKIGRRVTLVGLFTLGVAGVAYLF 349
 QY 418 FICVGRNVLTLLLFIAFAFISGGFQAAYVYTPVYPTATRALGLGTCGSMARVAGALITPF 477
 Db 350 LLAHDANTVLLTSALLSFPALLGAWGSLYATTPLEFPTPLRTTGMGLVSGVARLASVWSPS 409

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337
A:Accession: B83213
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <STO>
A:Cross-references: GB:AE004767; GB:AE004091; NID:g9949604; PIDN:AAG06855.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3467
C:Superfamily: hypothetical protein b1775

Query Match 16.4%; Score 466.5; DB 2; Length 455;
Best Local Similarity 27.7%; Pred. No. 1e-30; Indels 61; Gaps 13;
Matches 132; Conservative 94; Mismatches 189;
QY 72 DAVEAIGFGKFWKLSVLTGLAWMDAMEMMTLSILAPOLHCEWRLPSSQVALLTSVVFV 131
DB 10 ERLERPLSPYHRLVFVIALAFFDSDMLAMTFLGSIKAFFGLDSQAAGLASSSFF 69
QY 132 GMMSSSTINGNISDQYGRKTKGKIS-VLWTLTYGILSAPAPVYSWILVRLGLVGFIG-G 189
DB 70 GNVIGAAISGMLADRFGRKPVQASIVLWGLASYLCSTAGDLDLS-LTFYRVLLGICMGME 128
QY 190 VPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVFWLAVFVMPSLGWRWLLILSAVPLL 249
DB 129 FPTAQLSEMIPASRGKYIALMDGFPLGFVAACLSYFLPLTGTWRSIFLVALPAV 188
QY 250 LFAVLCFWLPESARYDVLSGNOEKATATUKRIATE-----NGAMPGLKLIISROEDR 302
DB 189 FVLAIRFLIPESRWLEQAGRRQADRVLRDIARVMRSGLGTLPPL-----RQPOR 242
QY 303 GKMR-----DLTPHFRWTLTLLWTFWESNAFSYVGLVLLTTLTFOAGDVCIGSSRK 354
DB 243 ERSRPGFFSAFELMSPAYRRTLTWGLFWFALLGFLGTLWSLALLQSGFA-----296
QY 355 KAVEAKCSLACEYLSSEDMYLLWTLSEFPGLVTLTWIIDRLGRKKTWALCF-----VI 409
DB 297 -----VTQSVYTYVL-ISLAGIPGFLCAAWLVESMGKRPSCVLMLLGGGAMA 342
QY 410 FSCSLLLFICVNRVTLTLF--IARAFISGGFQAAYVTPVRYTATRALGLGFCSCM 467
DB 343 YATGQTAVF---GGSLALLIGFLGAMQFFLFGWAVLYTYTPELYTSARATGSGFASAV 399
QY 468 ARVGALITPFIQAQVML-----ESSVYTLAVYSGCCLLAALASCLFPIETKGGGLQE 519
DB 400 GRIGSLGLPLVTGLVPLTGGQGVFTLGAL---CFGVAALVVMWAFGIETRGRTLEE 452

RESULT 8
JC4884
Organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
R:Accession: JC4884
R:Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.
Biochem. Biophys. Res. Commun. 224, 500-507, 1996
A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation transporter
A:Reference number: JC4884; MUID:96295517
A:Molecule type: mRNA
A:Residues: 1-593 <OKU>
A:Cross-references: DDBJ:D83044; NID:g1502282; PIDN:BAAL1754.1; PID:dl012421; PID:gl5022
A:Experimental source: kidney
C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 15.3%; Score 436; DB 2; Length 593;
Best Local Similarity 29.5%; Pred. No. 4.4e-28;
Matches 127; Conservative 77; Mismatches 148; Indels 78; Gaps 16;
QY 119 SNOVALTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAPAPVYSWILV 178

Db 146 SMLDLFOSVNVNGFFICAMMIGYLADRFGRKFCLLTILINAIISGALMAISPNTAMLV 205
QY 179 ---LRGLV---GFGIGGVQSVTLVYAEFLPMKARAKICILLIEVFWAIGTVFVFWLA--VF 230
DB 206 FRFLOGLVSKAGWLIGYI-----LITEFVGLGYRRMVGICYOIAFTVGL---LILAGVAY 257
QY 231 VMPSLGWRWLLTSLAVPULLLFAVLCFW-LPESARYDVLSGNOEKATATLKRITATENGAPM 289
DB 258 VIPN--WRWLFQAVTLPNFCF-LLYFWCIPESPRWLISONKIVKAMKILIKHIAKNGKSV 314
QY 290 PLGKLIISROEDRGK-----MRDLF-TPHFRWTLTLLWTFWESNAFSYVGLVLLTTELPO 343
DB 315 PVSLOQLTTPDEDAGKRLKPSILDVLRTPQIRKHTLILMYNWFSTSSVLYOGLIN---HMLG 371
QY 344 AGDVCGISRRKKAKEACSLACEYLSSEDMYLLWTLSEFPGLVTLTWIIDRLGRKKTW 403
DB 372 AGD-----NIYLDFFYSALVEFPFAAFIILITIDRGRRYPW 407
QY 404 ALC-FVIFSFCSLLLF1-----CVGRNVLTLTLFIARAFISGGFQAAYVYTPPE 450
DB 408 AVSNVAGAACLASVFIPTDLQWLKITTACLGRMGITM-----AYEMVCLVNAE 456
QY 451 VPTATRALGLGTCSGMARVGLITPFFIAQVWLESVYTLAVYSGCCLLAALASCLFPI 510
DB 457 LYPTTIRNLGLVLCSSMCDIGGIITPFLVYRLTDIWMPEPLVVFVAVVGLVAGLVLLLP- 515
QY 511 ETKGKLOQES 520
DB 516 ETKGKALPET 525

RESULT 9
S50862
Organic cation transporter protein OCT1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Nov-1999
R:Accession: S50862; S78533; I58089
R:Gruendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.
Nature 372, 549-552, 1994
A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.
A:Reference number: I58089; MUID:95082907
A:Accession: S50862
A:Molecule type: mRNA
A:Residues: 1-556 <GRU>
A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622
A:Experimental source: kidney
R:Gorboulev, V.G.
submitted to the EMBL Data Library, January 1995
A:Reference number: S78533
A:Accession: S78533
A:Molecule type: mRNA
A:Residues: 1-342, N',344-556 <GOR>
A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622
C:Keywords: glycoprotein; phosphoprotein; transmembrane protein
F:20-46/Domain: transmembrane #status predicted <TM1>
F:154-171/Domain: transmembrane #status predicted <TM2>
F:178-197/Domain: transmembrane #status predicted <TM3>
F:243-260/Domain: transmembrane #status predicted <TM4>
F:267-283/Domain: transmembrane #status predicted <TM5>
F:350-366/Domain: transmembrane #status predicted <TM6>
F:380-398/Domain: transmembrane #status predicted <TM7>
F:406-425/Domain: transmembrane #status predicted <TM8>
F:435-452/Domain: transmembrane #status predicted <TM9>
F:469-485/Domain: transmembrane #status predicted <TM10>
F:494-514/Domain: transmembrane #status predicted <TM11>
F:71,97,113,432/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:286,292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted
F:296,343,550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

Query Match 15.0%; Score 426.5; DB 2; Length 556;
Best Local Similarity 29.3%; Pred. No. 2.5e-27;

Db 121 LAUGALVPSAINLISEFAPAGRRSTLTVTWNSAFYSVGAVLSALLAIAMIPAMQWQSVFYV 180
QY 244 SAVPLLLFAVLCFWLPESARYDVLSCNQEKATATLKRIATENGAPMPLGKLIISROEDRG 303
Db 181 AVLPVLAVPLMLRWLPESAAFLLELKGRRAELDALLRKVDPDY-RPGAERANAVAAEAPSG 239
QY 304 KMRDLFTPHFRWTTLLWFIWFSNAPSYYYGLVLLTTTELFOAGDVCGIISRKKAVEAKCSL 363
Db 240 RVAQLFEGRQAVGTLLLLWAFAMCMLMSYGLNTWLPKLMAGGG----- 282
QY 364 ACEYLSSEEDYMDLWTTLTSEFFPGVLTWIIDRLGRKKTMAICFVIFSCSLLLLFICVGR 423
Db 283 ---YALGSSLAFVLTNVGATLGALEGGWLADRLGAGRTLVLFFALAAASLAALGLGPGP 339
QY 424 NVLTLLFIARAFISGGFOAAVYVTPETATRALGLGTCGSMARVGALITPFIQVVML 483
Db 340 WLLNGLLVVAGATTIGTLAVIHAYAAQFYPAWVRSTGVGWAAGVGRGAIAGPMLGGSLL 399

Search completed: March 13, 2002, 12:41:38
Job time: 84 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 17.05 Seconds
(without alignments)
1178.435 Million cell updates/sec

Title: US-09-911-667A-2
Perfect score: 2846
Sequence: 1 MEEDLFQLRLPVVKFRRTG.....MVGRMGAGVTRNSGSQ 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Sw:ssProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|--------------|--------------------|
| 1 | 1280 | 45.0 | 529 | Y0U1_CAEEL | P30638 caenorhabdi |
| 2 | 483 | 17.0 | 400 | YCE1_BACSU | O34691 bacillus su |
| 3 | 408 | 14.3 | 742 | SVY2_RAT | Q02563 rattus norv |
| 4 | 357.5 | 12.6 | 557 | 1 OCN2_HUMAN | O76082 homo sapien |
| 5 | 353 | 12.4 | 448 | 1 PCAK_PSEPU | O51955 pseudomonas |
| 6 | 346.5 | 12.2 | 457 | 1 PCAK_ACICA | Q43975 acinetobact |
| 7 | 339.5 | 11.9 | 557 | 1 OCN2_RAT | O70594 rattus norv |
| 8 | 332 | 11.7 | 557 | 1 OCN2_MOUSE | Q920e8 mus musculu |
| 9 | 322 | 11.3 | 451 | 1 YAJ_BACSU | P37514 bacillus su |
| 10 | 295.5 | 10.4 | 413 | 1 MUCK_ACICA | P94131 acinetobact |
| 11 | 295.5 | 10.4 | 466 | 1 BENK_ACICA | O30513 acinetobact |
| 12 | 286.5 | 10.1 | 443 | 1 YAAU_ECOLI | P31679 escherichia |
| 13 | 284 | 10.0 | 751 | 1 YLX5_CAEEL | P46501 caenorhabdi |
| 14 | 282.5 | 9.9 | 445 | 1 YGCS_ECOLI | Q46909 escherichia |
| 15 | 280.5 | 9.9 | 407 | 1 YB04_HAEIN | P71369 haemophilus |
| 16 | 279.5 | 9.8 | 403 | 1 MHPT_SCHPO | P77589 escherichia |
| 17 | 275 | 9.7 | 557 | 1 ITR2_SCHPO | P87110 schizosacch |
| 18 | 264.5 | 9.3 | 547 | 1 GRL1_LEIDO | O01440 leishmania |
| 19 | 262 | 9.2 | 522 | 1 STP1_ARATH | P23586 arabidopsis |
| 20 | 260.5 | 9.2 | 428 | 1 MMLH_ALCEU | O51798 alcaligenes |
| 21 | 255 | 9.0 | 541 | 1 HXT2_YEAST | P23585 schizosacch |
| 22 | 254 | 8.9 | 575 | 1 ITR1_SCHPO | Q10286 schizosacch |
| 23 | 249.5 | 8.8 | 473 | 1 PHDK_NOSCK | O24723 nocardioide |
| 24 | 249.5 | 8.8 | 546 | 1 HXT0_YEAST | P43581 saccharomyc |
| 25 | 243.5 | 8.6 | 461 | 1 CSBC_BACSU | P46333 bacillus su |
| 26 | 241.5 | 8.5 | 523 | 1 STC_RICCO | Q41144 ricinus com |
| 27 | 238.5 | 8.4 | 459 | 1 YDJK_ECOLI | P76230 escherichia |
| 28 | 234.5 | 8.2 | 559 | 1 YAEK_SCHPO | Q09852 schizosacch |
| 29 | 228 | 8.0 | 566 | 1 KHT2_KLULA | P53387 kluyveromyc |
| 30 | 227.5 | 8.0 | 587 | 1 PH84_YEAST | P25297 saccharomyc |
| 31 | 223 | 7.8 | 482 | 1 YFIG_BACSU | P54723 bacillus su |
| 32 | 221.5 | 7.8 | 464 | 1 GALP_ECOLI | P37021 escherichia |
| 33 | 221 | 7.8 | 570 | 1 HXT1_YEAST | P32465 saccharomyc |

ALIGNMENTS

| RESULT 1 | Y0U1_CAEEL | STANDARD; | PRT; | 529 AA. |
|----------|--|-----------|------|---------|
| ID | Y0U1_CAEEL | | | |
| AC | P30638; Q21101; | | | |
| DT | 01-APR-1993 (Rel. 25, Created) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | |
| DE | HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III. | | | |
| GN | ZK637.1 | | | |
| OS | Caenorhabditis elegans. | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; | | | |
| OC | Rhabditidae; Peloderinae; Caenorhabditis. | | | |
| OX | NCBI_TaxID=6239; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=BRISTOL N2; | | | |
| RX | MEDLINE=92168156; PubMed=1538779; | | | |
| RA | Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R., | | | |
| RA | Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A., | | | |
| RA | Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T., | | | |
| RA | Ainscough R., Waterston R.; | | | |
| RT | "The C. elegans genome sequencing project: a beginning."; | | | |
| RL | Nature 356:37-41(1992). | | | |
| RN | [2] | | | |
| RP | REVISIONS. | | | |
| RC | STRAIN=BRISTOL N2; | | | |
| RA | Durbin R.; | | | |
| RL | Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE). | | | |
| CC | -I- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; Z11115; CAA77460.1; | | | |
| DR | EMBL; Z22175; CAA77460.1; JOINED. | | | |
| DR | EMBL; Z22175; CAA80131.1; | | | |
| DR | EMBL; Z11115; CAA80131.1; JOINED. | | | |
| DR | PIR; S15786; S15786. | | | |
| DR | WormPep; ZK637.1; CE06638. | | | |
| DR | InterPro; IPR003662; sub_transportr. | | | |
| DR | Pfam; PF00083; sugar_tr; 1. | | | |
| DR | PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG. | | | |
| DR | PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG. | | | |
| KW | Hypothetical protein; Transmembrane; Transport. | | | |
| FT | TRANSMEM 86 106 POTENTIAL. | | | |
| FT | TRANSMEM 122 142 POTENTIAL. | | | |
| FT | TRANSMEM 158 178 POTENTIAL. | | | |
| FT | TRANSMEM 238 258 POTENTIAL. | | | |
| FT | TRANSMEM 320 340 POTENTIAL. | | | |
| FT | TRANSMEM 373 393 POTENTIAL. | | | |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 217 | 7.6 | 1222 | 1 | YMP3_CAEEL | Q10947 caenorhabdi |
| 35 | 216 | 7.6 | 468 | 1 | GLCP_SYNY3 | P15729 synechocyst |
| 36 | 216 | 7.6 | 496 | 1 | NANT_ECOLI | P41036 escherichia |
| 37 | 214.5 | 7.5 | 531 | 1 | GHT2_SCHPO | O74969 schizosacch |
| 38 | 214 | 7.5 | 567 | 1 | HXT3_YEAST | P32466 saccharomyc |
| 39 | 214 | 7.5 | 584 | 1 | ITR1_YEAST | P30605 saccharomyc |
| 40 | 213.5 | 7.5 | 592 | 1 | HXT5_YEAST | P38695 saccharomyc |
| 41 | 211 | 7.4 | 491 | 1 | XYLE_ECOLI | P09098 escherichia |
| 42 | 210 | 7.4 | 522 | 1 | STA_RICCO | Q10710 ricinus com |
| 43 | 210 | 7.4 | 535 | 1 | GHT6_SCHPO | O74849 schizosacch |
| 44 | 210 | 7.4 | 574 | 1 | GAL2_YEAST | P13181 saccharomyc |
| 45 | 208.5 | 7.3 | 452 | 1 | YDJE_ECOLI | P38055 escherichia |

```
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
SQ SEQUENCE 529 AA: 58317 MW: 8D2FF4CBA15ECD2D CRC64;

Query Match 45.0%; Score 1280; DB 1; Length 529;
Best Local Similarity 52.8%; Pred. No. 5.1e-84;
Matches 258; Conservative 74; Mismatches 125; Indels 32; Gaps 7;

Qy 42 HVG-----LEAVELDGAAPREFANPTDD-----TFWEDAVAEIGFGRFOWKLSVLT 90
Db 34 HVGDDFAVRYSNLDR-----ELGPTDQSPDSEKTTVDVEAEALGGRFQLKLSLT 89
Qy 91 GLAWMADAMEMMILSILAPOLHCEWRLPSQVALLTSVVFVGMSSSTLWGNISDOYGRK 150
Db 90 GMAWADAMEMMILSILPALACEMGSISSVQALVTCVFGSMMLSSSTFWGKICDRFGR 149
Qy 151 TGLKISVLTWLYGILSAFAPVYSWILVLRGLVGFVGIGVQSVTLVYAEFLPMKARAKCI 210
Db 150 KGLTFSTLVACIMGVISCMSPHFVYLLFFRGLTGFGIGVQSVTLVYAEFLPTAQAKCV 209
Qy 211 LLIEVFWAIGTVFVVLAVFVWPSLGRWMLLILSAVPLLLFAVLCFWLPESARYDVLSGN 270
Db 210 VLIESFWAIGVFEALLAYFVMSFGWRALMFLSLPLGIFAVASFWLPESARFDMASGH 269
Qy 271 QEKATATLKRIATENGAMPGLKLIISQ---EDRGKMRDLFTPHFRWTTLLLWIFWFSN 327
Db 270 PERALLETQAARMRVOLPTGRVLSSTKAGSESGDITANLLSPDLRKTITLLWCIWAIT 329
Qy 328 AFSYGLVLLTTELFOAGDVC---GISSRKKAWEAKCSLACEYLSSEEDYMDLLWTLSEFP 385
Db 330 AFSYGMVLTFTVLFQSHDECHGGLFSNGTOME-----VCQPLTRSDYFDLLSTLAEFP 384
Qy 386 GVLVTLWIIDRLGRKKTALCFVIFSCSLLLFCVGRNVLTLILFIARAFISGFGOAA 445
Db 385 GLIITVLIIEVGRKKTALAEYAFIETFLYCLDRFTVTVLIFARAFISGAFQAY 444
Qy 446 VYTPEVPTATRALGLGTCGMARVAGALITPFAQVLMLESSYVLPFLAVYSGCLLAALAS 505
Db 445 VYTPEVPTTLRAVGLGTCSAMARIGAIVAS-----EKSLSLPIGYGTAAIIGLIAS 497
Qy 506 CFLPIETKG 514
Db 498 LSLPIETKG 506

RESULT 2
YCEI_BACSU
ID YCEI_BACSU STANDARD; PRT; 400 AA.
AC Q34691;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YCEI.
OS YCEI.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB000617; BAA22256.1; -
DR EMBL; Z99105; CAB12089.1; -
DR Subtilist; BG12773; ycel.
DR InterPro; IPR003662; sub_transport.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 305 325 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
SQ SEQUENCE 400 AA: 43708 MW: E0AE0CEE5DD27395 CRC64;

Query Match 17.0%; Score 483; DB 1; Length 400;
Best Local Similarity 30.7%; Pred. No. 1.9e-27;
Matches 137; Conservative 75; Mismatches 160; Indels 74; Gaps 14;

Qy 83 QMKLSVLTGLAWMADAMEMMILSILAPOLHCEWRLPSQVALLTSVVFVGMSSSTLWGN 142
Db 9 QKLLGVAGLWLFDMVDVGLISFLIARLHVEWNLSPEMKWIGSVNSICMAAGAFLEGL 68
Qy 143 ISDOYGRKTGLKISVLTWLYGCI---LSAFAPVYSWILVLRGLVGFVGIGG-VQSVTLVY 198
Db 69 LADRIGRK---KVFIITLLCFSGISAFVTSLSAFLILREVIGMGLGGLPVAFTLS 125
Qy 199 EFLPMKARAKCILLIEVFWAIGTVFVVLAVFVWPSLGRWMLLILSAVPLLLFAVLCFWL 258
Db 126 EAVVPEKRGVIVLLEFSEWAVGNLAAALISYFVIFSGQAAALLTALTAFVLYLRTSL 185
Qy 259 PESARYDVLSGNQEKAIATLKRIATENGAMPGLKLIISQEDRGKMRDLFTPHFRWTTL 318
Db 186 PDSKYESLSAK-----KRSWMEN-----VKSVMARQYIR-----PTV 218
Qy 319 LLWFIWFSNAPSYGVLVLTTELFOAGDVCVGISSRKKAWEAKCSLACEYLSSEEDYMDLLW 378
Db 219 MLSIVMFCVVFYSYGMFLWPLSVMLLKGFMSITQS-----FEY-----VLL 258
Qy 379 TTLSEFPGLVLTWLIIDRLGRK-----KTMALCFVIFSCSLLLFCVGRNVLTL 430
Db 259 MTLAQLPGVFSAAWLEKAGRWILVVYLGITAGSAYFFGTADSLSLLTAG---VLLS 314
Qy 431 FTARAFISGGFOAAVYVTPETATRALGLGTCGMARVAGALITPFAQVLMLESSYVIT 490
Db 315 F----FNLGAWGLVYATPEQYPTAIRATGSGTTAAFGRIIGGIFGLVGLTAAARHISF 370
Qy 491 LAVYSGCLLAALA-SCFLPI--ETK 513
Db 371 V-IFSIFCIALLAVACILIMGKETK 395

RESULT 3
SVY2_RAT
ID SVY2_RAT STANDARD; PRT; 742 AA.
AC Q02563;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SYNAPTIC VESICLE PROTEIN 2 (SV2).
```

GN SV2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 CC [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
 RP TISSUE=Brain;
 RC MEDLINE=92390722; PubMed=1519064;
 RA Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.;
 RT "SV2, a brain synaptic vesicle protein homologous to bacterial
 RT transporters.";
 RL Science 257:1271-1273(1992).
 CC -!- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER
 CC TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.
 CC -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.
 CC -!- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND
 CC SPINAL CORD.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L05435; AAA2188.1; -.
 DR Pfam; PF00083; sugar_tr; 1.
 KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;
 KW Transmembrane.
 FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 164 182 POTENTIAL.
 FT DOMAIN 183 204 LUMENAL (POTENTIAL).
 FT TRANSMEM 205 225 POTENTIAL.
 FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 243 260 POTENTIAL.
 FT DOMAIN 261 262 LUMENAL (POTENTIAL).
 FT TRANSMEM 263 281 POTENTIAL.
 FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 295 322 POTENTIAL.
 FT DOMAIN 323 334 LUMENAL (POTENTIAL).
 FT TRANSMEM 335 355 POTENTIAL.
 FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 446 469 POTENTIAL.
 FT DOMAIN 470 592 LUMENAL (POTENTIAL).
 FT TRANSMEM 593 611 POTENTIAL.
 FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 627 647 POTENTIAL.
 FT DOMAIN 648 649 LUMENAL (POTENTIAL).
 FT TRANSMEM 650 669 POTENTIAL.
 FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 695 711 POTENTIAL.
 FT DOMAIN 712 712 LUMENAL (POTENTIAL).
 FT TRANSMEM 713 731 POTENTIAL.
 FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 742 AA; 82705 MW; 565DE7EF2929D5DB CRC64;

Query Match 14.3%; Score 408; DB 1; Length 742;
 Best Local Similarity 21.9%; Pred. No. 7.9e-22;
 Matches 151; Conservative 101; Mismatches 232; Indels 206; Gaps 20;
 QY 20 GESARSED-----DTASG--EHEVQIEGVHVGLEAVE-----LD 51
 Db 68 GEGAQDDEEGCASSDTEGDEHDEDEIYEYQGIIPRAESGKGGERMADGAPLAVRGSL 127
 QY 52 DGAAVP-----KEFANPTDDTFMVEDAVEAIGFKFOWKLSVITGLAWMADAM 99
 Db 128 DGEPPGGGRGEAQRKRDRELAQ-----QYETILRCGHGFWTLYFVLGLALMADGV 181

QY 100 EBMILSLILAPOLHCEWRLPSPHOVALLTSYVFVGMSSSTLNGNISDOYGRKTKGLKISVLW 159
 Db 182 EVFVGVFLPSAEKDMCLSDSNKGMGLIYVYLMVGAFLMGLADRLGRROCLLSLSV 241
 QY 160 TLXYGILSAPAPVYSWILVRLGLVGFEGIGG-VPOSVTLYAEFLPMKARAKCILLIEVFWA 218
 Db 242 NSVPFAFFSSVQGYGTFLFCRLLSGVGLGGSIPIVFSVSEFLAQEKRGHLSWLCMFWM 301
 QY 219 IGTVEFVVLAVFVMPISLG-----WRMLLILSAPVPLLLFAVLCWLPESARYD 265
 Db 302 IGGVYAAAMAWAIIIPHYGWSFGMSAYQFHSRVFVLVFAFPVSFAIGALTQTPESPRFF 361
 QY 266 VLSNQEKAIATLARIATEN---GAP---MPLCKLIISROEDR----- 302
 Db 362 LENGKHDEAMVVLKQVHDTNMRAGHPERVSFVTHIKTIHQEDELIEIOSDTGTWYQRMG 421
 QY 303 -----GKMRDLTPHFRWTTLLWLFNFSNFSYVYGLV----- 336
 Db 422 VRALSLGGQVWGNFLSCFSPFYRRITLMMGMVMTMSYVYGLTWPPDMIRHLQAVDYA 481
 QY 337 -----LTELFOAGDVCV---ISSRKAVERAKCSL----- 363
 Db 482 ARTKVFCGERVEHVTFTLENQIHRGGQYFNDKFIQURLKSVSFEDSLFECYFEDVTS 541
 QY 364 -----ACEYLSSEEDY-----MDLLWT----- 379
 Db 542 SNTFFRNCFTFNTFYNTDLEFYKFNVSRLVNSTFLHNKGCPLDVTGTGEGAYMVYFVS 601
 QY 380 ---TLSEFPFGVLVTLWIIDRLGRKKTMCALCFVTFSCSLLLFICVGRN---VLTLLLFIA 433
 Db 602 FLGTLAVLPGNIVSALLMDKIGRLMLAGSSVL--SCVSCFFLSFGNSESAMIALCLCFG 659
 QY 434 RAFTSGGFQAAVYVTPVYPTATRALGIGTCSGMARVACALI-----TPFIAQVMLESSVY 488
 Db 660 GVSIA-SWNALDVLTVELYPSDKRTTARFELNACKLAALVIGISIFTSFVGITKAAPILF 718
 QY 489 LTLAVYSGCCLLAALASCLFLPIETKGGGLQ 518
 Db 719 ASAALGSSLALKLP-----ETRGQVLO 742
 RESULT 4
 OCN2_HUMAN
 ID OCN2_HUMAN STANDARD; PRT; 557 AA.
 AC 076082;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 2,
 DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).
 GN SLC22A5 OR OCTN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=98289574; PubMed=9618255;
 RA Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;
 RT "CDNA sequence, transport function, and genomic organization of human
 RT OCTN2, a new member of the organic cation transporter family.";
 RL Biochem. Biophys. Res. Commun. 246:589-595(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=98352077; PubMed=9685390;
 RA Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M.,
 RA Sai Y., Tsuji A.;
 RT "Molecular and functional identification of sodium ion-dependent, high
 RT affinity human carnitine transporter OCTN2.";
 RL J. Biol. Chem. 273:20378-20382(1998).
 RN [3]

RP SEQUENCE FROM N.A.
RX MEDLINE=99113835; PubMed=9916797;
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,
RA Nikaide H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T.,
RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,
RA Tsuji A.;
RT "Primary systemic carnitine deficiency is caused by mutations in a
RT gene encoding sodium ion-dependent carnitine transporter.";
RL Nat. Genet. 21:91-94(1999).
RN [4].
RP CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter.";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
RN [5].
RP VARIANT CDSP GLN-169.
RX MEDLINE=99355597; PubMed=10425211;
RA Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K.,
RA Gerbitz K.-D., Killmann M.W.;
RT "Carnitine transporter OCTN2 mutations in systemic primary carnitine
RT deficiency: a novel Arg169Gln mutation and a recurrent Arg282Ser
RT mutation associated with an unconventional splicing abnormality.";
RL Biochem. Biophys. Res. Commun. 261:484-487(1999).
RN [6].
RP VARIANT CDSP CYS-211.
RX MEDLINE=99408248; PubMed=10480371;
RA Vaz F.M., Scholte H.R., Ruiter J., Hussaarts-Odiijk L.M.,
RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,
RA Wanders R.J.A.;
RT "Identification of two novel mutations in OCTN2 of three patients with
RT systemic carnitine deficiency.";
RL Hum. Genet. 105:157-161(1999).
RN [7].
RP VARIANT CDSP LEU-478.
RX MEDLINE=99172075; PubMed=10072434;
RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,
RA Wanders R.J., Fok T.F., Hjelm N.M.;
RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to
RT deficient cellular carnitine uptake in primary carnitine deficiency.";
RL Hum. Mol. Genet. 8:655-660(1999).
RN [8].
RP CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.
RX MEDLINE=20026865; PubMed=10559218;
RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;
RT "Mutations in novel organic cation transporter (OCTN2), an organic
RT cation/carnitine transporter, with differential effects on the
RT organic cation transport function and the carnitine transport
RT function.";
RL J. Biol. Chem. 274:33388-33392(1999).
RN [9].
RP VARIANTS CDSP ARG-283 AND PHE-446.
RX MEDLINE=20081068; PubMed=10612840;
RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,
RA Tsuji A.;
RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a
RT patient with primary systemic carnitine deficiency.";
RL Hum. Mutat. 15:118-118(2000).
RN [10].
RP VARIANT CDSP LYS-452.
RX MEDLINE=20145665; PubMed=10679939;
RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;
RT "A missense mutation in the OCTN2 gene associated with residual
RT carnitine transport activity.";
RL Hum. Mutat. 15:238-245(2000).
RN [11].
RP FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC [12].
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC [13].
CC TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,
CC HEART AND PLACENTA.
CC [14].
CC DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY
CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE
CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE
CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND
CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL
CC MYOPATHY OR CARDIOMYOPATHY.
CC [15].
CC SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
CC [16].
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC [17].
CC EMBL: AF057164; AAC24828.1; -
CC EMBL: AB015050; BAA29023.1; -
CC EMBL: AB016625; BAA36712.1; -
CC MIM: 603377; -
CC MIM: 212140; -
CC InterPro: IPR003662; sub_trnsport.
CC Pfam: PF00083; sugar_tr; 1.
CC PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
CC Transmembrane; Glycoprotein; Disease mutation.
CC TRANSPORT: Transmembrane; Glycoprotein; Disease mutation.
CC TRANSMEM 21 41 POTENTIAL.
CC TRANSMEM 143 163 POTENTIAL.
CC TRANSMEM 173 193 POTENTIAL.
CC TRANSMEM 196 216 POTENTIAL.
CC TRANSMEM 233 253 POTENTIAL.
CC TRANSMEM 258 278 POTENTIAL.
CC TRANSMEM 343 363 POTENTIAL.
CC TRANSMEM 372 392 POTENTIAL.
CC TRANSMEM 414 434 POTENTIAL.
CC TRANSMEM 437 457 POTENTIAL.
CC TRANSMEM 489 509 POTENTIAL.
CC CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).
CC VARIANT 169 169 R -> Q (IN CDSP).
CC VARIANT 211 211 /FTid=VAR_009252.
CC VARIANT 211 211 Y -> C (IN CDSP).
CC VARIANT 283 283 /FTid=VAR_009253.
CC VARIANT 283 283 W -> R (IN CDSP).
CC VARIANT 446 446 /FTid=VAR_009254.
CC VARIANT 446 446 V -> F (IN CDSP).
CC VARIANT 452 452 /FTid=VAR_009255.
CC VARIANT 452 452 E -> K (IN CDSP).
CC VARIANT 478 478 /FTid=VAR_009256.
CC VARIANT 478 478 P -> L (IN CDSP: LOSS OF CARNITINE
CC TRANSPORT BUT STIMULATED ORGANIC CATION
CC TRANSPORT).
CC MUTAGEN 352 352 /FTid=VAR_009257.
CC MUTAGEN 352 352 M -> R: LOSS OF BOTH CARNITINE AND ORGANIC
CC CATION TRANSPORT FUNCTIONALITIES.
CC SEQUENCE 557 AA; 62751 MW; 928B1F6FEFF63C48D CRC64;

Query Match 12.6%; Score 357.5; DB 1; Length 557;
Best Local Similarity 24.8%; Pred. No. 2.3e-18;
Matches 124; Conservative 96; Mismatches 204; Indels 77; Gaps 16;
Qy 48 VELDDGAAPV---REFANPTDDTFMVEDAVEAIG-----FGKFWKLSVLTGLAMMA 96
Db 69 LRLRDGPVPHSCRRYRLATIANF-----SALGREGVDRVDLGOLEOE-SCLDGWESQ 121
Qy 97 DAMEMMLISLAPQLHCEWRL--PSQVALLTSVVFVGMSSSTLWGNISDQYGRKTGL 153
Db 122 DVLSTIVT-----EWNLCEDDWKAPLTISLFFVGVLLGSFISQSLSDRGRKNVL 173

```
QY 154 KISVLWTLYGILSAPAPYVSWILVRLGVFG-IGVPOSVTLYAEFLPMKARAKCILL 212
DB 174 FVTMGQTFGSFLQIFSKNFEMFVLVLMGQOISNYAAAFVLGTEILGKSVRIIFSTL 233
QY 213 -IEVFWAIGVFEVWLAFAVMPVSLGNRWLLILSAVLLFVAVLCFWLPESARYDVLGSGNQ 271
DB 234 GVCIFAAGVWVLPFLPAYFIR---DWRMLLVATMGVLCVALWFIPEPRLISQGRF 290
QY 272 EKAIALTKRIATENGAMP-----LGLIISROEDRGKMRDLFTPHFRWTTLLLWFI 323
DB 291 EAEVIRKAANGANGVWPSTIFDPSELQDLSKKQOOSHNLILLRTWNRIMVTINSIML 350
QY 324 WFSNARSYYGLVLTTLFELFOAGDVCGISSRKAKEAKCSLACEYLSEEDYMDLLTTLSE 383
DB 351 WMTISVGYFGLSLDTNLH--GDI-----FVNCFLSAMVE 383
QY 384 FPGVLVTLWIIDRLGRKKTWALCFVIFSCSLLLIFCVGRNVL---TLLFTARAFISG 439
DB 384 VPAYVALLWLLQVLPRIYSNAT--ALFLGGSVLLFMQLVPPDLYLATVLMVYKFGVTA 441
QY 440 GFOAAVYVTPVPTATRALGLTSCGMARGALITPFTIAQVMLESSVYLTAVYSGCCL 499
DB 442 AFSNVVYVTAELPTVVRNMGVGSVSATSLGSIILSPYFVYLCAVDR-FLPYILMGSLTI 500
QY 500 LAALASCFPIETKGGGLQES 520
DB 501 LTAILTLFLP-ESFGTPLPDT 520

RESULT 5
PCAK_PSEPU STANDARD; PRT; 448 AA.
AC Q51955;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 4-HYDROXYBENZOATE TRANSPORTER.
GN PCAK.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae.
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PRS2000;
RX MEDLINE=95050205; PubMed=7961399;
RA Harwood C.S., Nichols N.N., Kim M.-K., Ditty J.L., Parales R.E.;
RT "Identification of the pcarkF gene cluster from Pseudomonas putida:
RT involvement in chemotaxis, biodegradation, and transport of
RT 4-hydroxybenzoate.";
RL J. Bacteriol. 176:6479-6488(1994).
CC -!- FUNCTION: TRANSPORTER FOR 4-HYDROXYBENZOATE. ALSO REQUIRED FOR
CC CHEMOTAXIS TO AROMATIC ACIDS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U01895; AAA85137.1; -.
CC InterPro: IPR003662; sub.transprot.
CC Pfam; PF00083; sugar.tr.1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
```

```
FT TRANSMEM 95 115 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
SQ SEQUENCE 448 AA; 47176 MW; 6D51C143123E99BC CRC64;

Query Match 12.4%; Score 353; DB 1; Length 448;
Best Local Similarity 26.6%; Pred. No. 3,9e-18;
Matches 122; Conservative 77; Mismatches 189; Indels 70; Gaps 14;

QY 62 NPDDTFMVEDAVEAIGFGKFWKLSVLTGLANMADAMEMMILSILAPOLHCEWRLPSWQ 121
DB 6 NSVGKSLDVGSFINQQPLSRIOVRVLLCLFLVFLDGLDTAAMGFIAPALSQEWGIDRAS 65
QY 122 VALLTSVYFVGMSSSTLWGNISDQYGRKTKLSVLTWLYGILSAFAPYVSWILVRLG 181
DB 66 LGPMSAALIGWVFGALSGPLADRFGRKGVLVGAVLVFGGFSLASAVATNVLDQLVLR 125
QY 182 LVGFGIG-GVPOSVTLYAEFLPMKARAKCILLIEVF--WAIGTVFEVVLAVFVMPVSLGWR 238
DB 126 LTGLGLGAGPNATLLSEYTP--ERLKSLLVTSFMFCFNLGMAGGGFISAKMIPAYGWH 183
QY 239 WLLILSAV-PLLLFAVLCFWLPESARYDVL--SGNQEKAIATLKRKIATENGAPMLGLKLI 296
DB 184 SLLVIGVLPALLALVLMWLPESARELVVRNRTDKIRKTLSPVIAQVVA--EAGSFSV 241
QY 297 SROE---DRGKMRDLTPHERWTTLLLWFIWFSNAFYVGLVLTTELFOAGDVCGISSR 353
DB 242 PEQKAVAA-RSVFAV:PSGTVGLGTMLLWLYFMGLVIVY---LLTSWLTPLMRDQSGASME 298
QY 354 KKAWEAKCSLACEYLSSEEDYMDLLTTLSEPPGVLVTL---WIIDRLGRKKTALCFV-- 408
DB 299 QAFAFG-----ALFQFGVLSAVGVGMWAMDRYNPHKVIIGIFVLLA 338
QY 409 -IFSFCSLLLFICVGRN-----VLTLLLFARAFISGFOAAVYVTPVPTATRALGLG 462
DB 339 GVFAV-----AVGQSLGNITVLTATLVLAGMVCVNAQSAMPVLAARFPTQGRATGVS 391
QY 463 TCSGMARVAGALITPFTIAQVMLESSVYLTAVYSGCCLL 500
DB 392 WMLGIGRFGAI-----LGAWSGATLL 412

RESULT 6
PCAK_ACICA STANDARD; PRT; 457 AA.
AC Q43975;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 4-HYDROXYBENZOATE TRANSPORTER.
GN PCAK.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BD413 / ADP1;
RX MEDLINE=94341565; PubMed=8063101;
RA Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.-L.,
RA Ornstom L.N.;
RT "Contrasting patterns of evolutionary divergence within the
RT Acinetobacter calcoaceticus pca operon.";
RL Gene 146:23-30(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
```

```
CC CC (POTENTIAL).
CC CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; L05770; AAC37151.1; -.
DR DR InterPro; IPR003662; sub_trnsporttr.
DR DR Pfam; PF00083; sugar_tr; 1.
DR DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transport; Transmembrane; Inner membrane.
FT TRANSMEM 35 55 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 102 122 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 275 295 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 339 359 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
FT TRANSMEM 401 421 POTENTIAL.
FT TRANSMEM 427 447 POTENTIAL.
SQ SEQUENCE 457 AA; 49277 MW; 4F5B5F77361A1567 CRC64;

Query Match 12.2%; Score 346.5; DB 1; Length 457;
Best Local Similarity 24.9%; Pred. No. 1.le-17;
Matches 115; Conservative 92; Mismatches 197; Indels 57; Gaps 14;

QY 56 VPKEFANPTDDTFWE---DAVEAIG---FGKFWKLSVLTGLAWMADAMEMMTLSLA 108
DB 1 MPKE-ANMASQDYATQSRSLDAQALINDAPLSYQWLAIYVCFLLVFDGIDTAMAGFIA 59

QY 109 POLHCEWRLPSQVALLTSVYFVGMMSSSTLWGNISDQYGRKTGLKISVLWTLTYGILSA 168
DB 60 PALAQDWGVDRSQLGPNVSAALGGMIIGALVSGPTADRFGRKIVLSMSMLVFGGFTLACA 119

QY 169 PAPVYSWLVLRLGLVFGFVG-GVPOSVTLYAEEFLPKAKAKICILLIEFVFAIGTVFEVL 227
DB 120 YSTNLDLSLVIFRFLTGIGLGAAMPNATTLFSEYCPARISLLVTCMFCGYNLGMAIGFI 179

QY 228 AVFVMPSLGWRWLLLSA-VPLLFAVLFCFPLPESARYDVLSG-NOEKAIATLKRIA--- 282
DB 180 SSWLIPAFGWHSLFLGCGWAPLMLLVIFLFPESYRELIVKGNKTKVKRQILSRIPQK 239

QY 283 ----TENGAPMLPLKLIISROEDRGMKRDLLTPHFRWTLTLLWFTWFSNARSYYGLVLT 338
DB 240 VOGVTEFHVP-----BEKVEAGTKKGVFGVSAKYVKGTVLLWTVYFGLVMYI---LLT 292

QY 339 TELFQAGDVGCGTSSRRKKAWEAKCSLACEVLSBEDYMDLLWTLTSLFPQGVLT---WIID 395
DB 293 SWLPLTMRGTASLERAA-----FLG-----GLFQGGVSLNLFICWAND 332

QY 396 RLGRKKTALCEVIFPFCSLLLFCVGRN-----VLTLLFTIARFISGFOAAVYVYPE 450
DB 333 RENPNRIIA---GYLAAGIFAVTVGQSLNSPTLLALFILCAGIAGVANGAQSMPVLSAR 388

QY 451 VYPTATRALGLTGCSMARVGLIIPFLAQVMLESSVYLT 491
DB 389 FYPTOCATGVANMSGIGRFAGVAGVAGVILGNNSFTM 429

RESULT 7
OCN2_RAT
ID OCN2_RAT STANDARD; PRT; 557 AA.
AC O70594; Q9QWL0;
```

```
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)
DE (UST2R) (CTI).
GN SLC22A5 OR OCTN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=98200080; PubMed=9541011;
RA Schoenig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,
RA Gruendemann D.;
RT "Molecular cloning and characterization of two novel transport
RT proteins from rat kidney.";
RT FEBS Lett. 425:79-86(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
RX MEDLINE=99011422; PubMed=9792817;
RA Sekine T., Kusuhashi H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
RA Kanai Y., Endou H.;
RT "Molecular cloning and characterization of high-affinity carnitine
RT transporter from rat intestine.";
RT Biochem. Biophys. Res. Commun. 251:586-591(1998).
RN [3]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter";
RT J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES
CC AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND
CC ARTERIOLES IN THE HEART, IN THE LABIRINTHINE LAYER OF THE
CC PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE
CC BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; AJ001933; CAA05106.1; -.
DR EMBL; AB017260; BAA34399.1; -.
DR EMBL; AF110416; AAB54059.1; -.
DR InterPro; IPR003662; sub_trnsporttr.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
```


[2] SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.
RN
RP STRAIN-C3H;
RC MEDLINE-99057546; PubMed-9837751;
RX Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;
RA "A missense mutation of mouse OCTN2, a sodium-dependent carnitine
RT cotransporter, in the juvenile visceral steatosis mouse."; [3]
RL Biochem. Biophys. Res. Commun. 252:590-594(1998).
RN

WU X., HUANG W., PRASAD P.D., SETH P., RAJAN D.P., LEIBACH F.H., CHEN J., CONWAY S.J., GANAPATHY V.: "Functional characteristics and tissue distribution pattern of organic

CC transporter; ;
CC J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL
CC STEATOSIS (JVS).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

| | | | | |
|----------|---------|------------------------|------------------|--------|
| TRANSMEM | 233 | POTENTIAL. | | |
| FT | 238 | POTENTIAL. | | |
| TRANSMEM | 258 | POTENTIAL. | | |
| FT | 342 | POTENTIAL. | | |
| TRANSMEM | 374 | POTENTIAL. | | |
| FT | 427 | POTENTIAL. | | |
| TRANSMEM | 401 | POTENTIAL. | | |
| FT | 451 | POTENTIAL. | | |
| TRANSMEM | 431 | POTENTIAL. | | |
| FT | 489 | POTENTIAL. | | |
| TRANSMEM | 57 | POTENTIAL. | | |
| FT | 57 | POTENTIAL. | | |
| CARBOHYD | 64 | N-LINKED (GLCNAC. . .) | (POTENTIAL). | |
| FT | 64 | N-LINKED (GLCNAC. . .) | (POTENTIAL). | |
| CARBOHYD | 91 | N-LINKED (GLCNAC. . .) | (POTENTIAL). | |
| FT | 91 | N-LINKED (GLCNAC. . .) | (POTENTIAL). | |
| CARBOHYD | 322 | N-LINKED (GLCNAC. . .) | (POTENTIAL). | |
| FT | 322 | N-LINKED (GLCNAC. . .) | (POTENTIAL). | |
| VARIANT | 352 | L -> R (IN JVS) | | |
| SQ | 557 AA; | 62779 MW; | 6093F0EE9612B204 | CRC64; |

| | | |
|----|---|--|
| RN | [1] | |
| RC | SEQUENCE FROM N.A. | |
| RC | STRAIN=C57BL/6J; TISSUE=Kidney; | |
| RX | MEDLINE=J9113835; PubMed=9916797; | |
| RA | Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N., | |
| RA | Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T., | |
| RA | Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M., | |
| RA | Tsuji A. | |
| RT | "Primary systemic carnitine deficiency is caused by mutations in a | |
| RT | gene encoding sodium ion-dependent carnitine transporter." | |
| RL | Nat. Genet. 21:91-94(1999). | |

[1]
SEQUENCE FROM N.A.
STRAIN=BD413 / ADP1;
MEDLINE=97440148; PubMed=9294456;
Collier L.S., Nichols N.N., Neidle E.L.;
"benk encodes a hydrophobic permease-like protein involved in benzoate
degradation by Acinetobacter sp. strain ADP1.";
J. Bacteriol. 179:5943-5946(1997).
CC -!- FUNCTION: PROBABLE UPTAKE OF BENZOATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF009224; AAC46425.1; .
CC InterPro: IPR003662; sub_transporter.
CC Pfam; PF00083; sugar_tr; 1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Transport; Transmembrane; Inner membrane.
CC FT TRANSMEM 23 43 POTENTIAL.
CC FT TRANSMEM 61 81 POTENTIAL.
CC FT TRANSMEM 94 114 POTENTIAL.
CC FT TRANSMEM 120 140 POTENTIAL.
CC FT TRANSMEM 151 171 POTENTIAL.
CC FT TRANSMEM 182 202 POTENTIAL.
CC FT TRANSMEM 259 279 POTENTIAL.
CC FT TRANSMEM 298 318 POTENTIAL.
CC FT TRANSMEM 326 346 POTENTIAL.
CC FT TRANSMEM 349 369 POTENTIAL.
CC FT TRANSMEM 388 408 POTENTIAL.
CC FT TRANSMEM 415 435 POTENTIAL.
CC SEQUENCE 466 AA; 51403 MW; D60C2E61D512F334 CRC64;
Query Match 10.4%; Score 295.5; DB 1; Length 466;
Best Local Similarity 23.0%; Pred. No. 5e-14;
Matches 104; Conservative 82; Mismatches 179; Indels 87; Gaps 12;

or send an email to license@isb-sib.ch).
CC EMBL; U87258; AAC27117.1; .
CC InterPro: IPR003662; sub_transporter.
CC Pfam; PF00083; sugar_tr; 1.
CC DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Transport; Transmembrane; Inner membrane.
CC FT TRANSMEM 17 37 POTENTIAL.
CC FT TRANSMEM 54 74 POTENTIAL.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 113 133 POTENTIAL.
CC FT TRANSMEM 146 166 POTENTIAL.
CC FT TRANSMEM 172 192 POTENTIAL.
CC FT TRANSMEM 229 249 POTENTIAL.
CC FT TRANSMEM 267 287 POTENTIAL.
CC FT TRANSMEM 294 314 POTENTIAL.
CC FT TRANSMEM 319 339 POTENTIAL.
CC FT TRANSMEM 362 382 POTENTIAL.
CC FT TRANSMEM 383 403 POTENTIAL.
CC SEQUENCE 413 AA; 45245 MW; 2D88CE31C4C5CC65 CRC64;
Query Match 10.4%; Score 295.5; DB 1; Length 413;
Best Local Similarity 24.9%; Pred. No. 4.4e-14;
Matches 111; Conservative 80; Mismatches 179; Indels 75; Gaps 18;
Qy 80 GKFWKLSVLTG-LAWMADAMEMMILSLAPQLHCEWRLPSQVALLTSVVFVGMSSST 138
Db 11 GSHTWKIAFLAFALVDGADLMLLSYLSNSIKAEFNLSLTVAGMLGFTLAGMAIGGI 70
Qy 139 LWNISDQYGRKTKLSVLTWLYYGLIS-AFAPVYSWIL-VLRGLVGFVGIGVQSV- 194
Db 71 FCGWACDRFGR--VRIWVISILTSILTCGLTQSFIOFQVLRFPFASLGLSLYIACN 127
Qy 195 TLXAEFLPMKARAKILLIEVFWAIGTVFVAVFVAVFVAVFVAVFVAVFVAVFVAVL 254
Db 128 TLMAYVPTKYRTVLTGLQAGTVGVIVATLLAGLIPDHGVRVLYVAVIIPVLMVLM 187
Qy 255 CFWLPESARYDVLSCNQEKAIATLKRIATENGAMPGLKLIISROEDRGKMRDLFTPHR 314
Db 188 HFFVPEPAAM-----QOSRLAPSKQFETVKTSAF-----KLIFQDKRNRN----- 227
Qy 315 WTTLLWFIFNSNA----FSYGLVGLVLTTELFOAGDVCIGSSRRKKAWEAKCSLACEYISE 370
Db 228 -----MFLMALTAGFLQFGYGVNNMMPVLES-----ELCMKEKEM 265
Qy 371 EDYMDLLWTLSEPPGVLTWLIIDRLGRKKTALCEV-IFSFCSLLLFICVGRNVLTLL 429
Db 266 TAYMVGYTAN--ILGKILAGFMADKIGRRFTYAFAIGAIGTAIFLPLIVFYNSPNILYLL 323
Qy 430 LFIARAF SG---GFOAAVYTPVYPTATRALGLGTCGSMARVAGALITP----FIAQVM 482
Db 324 VIF--GFLYGIPIGVNATYM--TESFTAIRGTAIGAYNVGRGLGAIAIAPATIGFLAS-- 377
Qy 483 LESSVYLT-----AVYSGCCLLAL 503
Db 378 -GGSIGLGFVVMGAAYFICGVIPAL 401
RESULT 11
BENK_ACICA STANDARD; PRT; 466 AA.
ID BENK_ACICA
AC O30513;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BENZOATE TRANSPORT PROTEIN.
GN BENK.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=471;

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; U12965; AAA20607.1; --
CC WormPep; F23F12.5; CE01252.
CC InterPro; IPR002184; Strb.
CC Pfam; PF02175; Strb; 1.
CC Pfam; PF00083; sugar.tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 34 44 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 102 122 POTENTIAL.
FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 515 535 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
FT TRANSMEM 583 603 POTENTIAL.
FT TRANSMEM 614 634 POTENTIAL.
FT TRANSMEM 678 698 POTENTIAL.
SQ SEQUENCE 751 AA; 84832 MW; A6C4F43540295EFC CRC64;

Query Match 10.0%; Score 284; DB 1; Length 751;
Best Local Similarity 25.1%; Pred. No. 5, 3e-13;
Matches 111; Conservative 82; Mismatches 149; Indels 100; Gaps 24;

QY 119 SNOVALLTSVVFVGMSSSTLMGNTSDQYGRKT----GLKISVLWPLYGILSAPFVYS 174
DB 316 AYDAAVATIQFIVGLGALTYGHLGHDFGRKPVSEFGISVGLI---PGVASGFAPSWE 371
QY 175 WILVLRGLVCGFVGIGVQSVTLYAELPMKARAKCILLIEVF-WAIGTVFVVVLAFFVMP 233
DB 372 VFAAFRPIVGTSTASI--LIVFYAYILEFIEPQRFVLSFFNWGYARLV-FTLACFICG 428
QY 234 SLGWR-----WLLLSAVPLLLFAVLCFWLPESARYDVLSCNGOEKAATLKRATFENGA 287
DB 429 Y--WRSAATISLLSLPILVLLI-----LPSPKAWNTKKRFRDARAARVAMLSGI 480
QY 288 P-----MPLGKLLISQEDRGKRDLTFTPHRWT----TLLLFIFWFSNAFSYGLV 335
DB 481 PYVNDEDSQISEIKLEEKSTKIYTKMDLFT---SWTIAYRTIVGSLMFSTLSAFGSD 537
QY 336 LLTTELFQAGDVCGVGISSRRKKAVERKSLACEYLSEEDYMDLLTTLSEFPGLVLTWIID 395
DB 538 L-----NSGNLAGNF-----YLSQ-----FVSGAVTAFAKIFV--FLLD 569
QY 396 -----RLGRRKKTALCFVIFSCSLLLFICV-----GRVLLTLL--FTARAF 436
DB 570 TYVPSEFDRRLRHQYQIAM---ILCYVIMVILPESDGSQSGSRDLAIHINIIGVSF 626
QY 437 ISGFGQAAVYVTEVYPTATRALGLGTCGSMARVGALITPFAQVMLESSVYLTAVYSG 496
DB 627 IEITWDACYLVAVECFPTKRTIGTCTSLARTGALLAPQAMYL---SDIYRP-APYAV 682
QY 497 CC---LALALASC-FLPIETKG 514
DB 683 VCSIGTISLLISCVFLP-DPKG 703

RESULT 14
YGS_ECOLI
ID YGS_ECOLI STANDARD; PRT; 445 AA.
AC Q46909;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YGCS.
GN YGCS OR B2771.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U29579; AAA69281.1; ALT_INIT.
DR EMBL; AE000360; AAC75813.1; ALT_INIT.
DR EcoGene; EG13126; YGCS.
DR InterPro; IPR003662; sub.trnsportr.
DR Pfam; PF00083; sugar.tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 287 307 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 370 390 POTENTIAL.
FT TRANSMEM 401 421 POTENTIAL.
SQ SEQUENCE 445 AA; 48234 MW; B59E452721B15774 CRC64;

Query Match 9.9%; Score 282.5; DB 1; Length 445;
Best Local Similarity 24.0%; Pred. No. 4e-13;
Matches 114; Conservative 82; Mismatches 212; Indels 67; Gaps 15;

QY 74 VEAIGFKTQWKLSVLTGLAWMADAMMTLSILAPOLHCENRPLSQVQVALLTSVVFVGM 133
DB 8 MDOLPLNRHCRIAALTFGAHLTDGVVLGVIGYAIQLTPAMQLTPFMAGMIGSALLGL 67
QY 134 MSSSTLWGNISDOYGRKTGLKISVLWTLYYGILSAPFVYSWILVLRGLVGFVGIGVQPS 193
DB 68 FGLSVLWGNISDHIGRQKIFTFSLITLASFLQFFATTFEHLIGLILIGLGG-DYS 126
QY 194 V--TLVAEFLPMKARAKCILLIEVFWAIGTVFVLAFFVMPSL-----GHRWL 240
DB 127 VGHLLAEFSRRHRGILLGAFSVVMTVG-----YVLASTAGHHFISENPEARWL 177
QY 241 LILSAVPLLLFAVLCFWLPESARYDVLSCNGOEKAATLKRATENGAPMLGKLIISROE 300
DB 178 LASAALPALLITLLRWGTPEPSRWLRQGRFAEHAHVHRFY---GPHVLLGDVGVFATH 234
QY 301 DRGKMRDLFTPHFRWTLTLWFIFWFSNAFSYVGLVLTTLTTELFQAGDVCGVGISSRRKKAVERK 360

```
Db 235 KH--IKTFSSRY-----WRTAFNSVFFVCLVIPWF-----VIYTW 269
QY 361 CSLACEYLSEEDYM--DLLWTTLSEPPGVLTWIIIDRLGRKKTALCFVIFSCSLLLF 418
Db 270 LPTIAOTIGLEDALTASMLNALL--IVGALLGLVLTHLAHRKFLUGLSLLLA--ATLVVM 327
QY 419 ICV--GRNVLTLLIFIA--RAFISGGFOAAVYVTPETATRALGLGTCGSMARVGAII--T 475
Db 328 ACLPGSSSLTLLLVLFSTTISAVNSLVGLIPAESPTDIRSLGVGFATAMSRGLRAVST 387
QY 476 PFTAQVWLESSVYLTAVYSGCCLLAALASCFPIETKG-----GGLQSS 521
Db 388 GLLPWLAQMGQVTLTLLATVLLGVFVVTWLPAPETKALPLVAAGNVGAGNERS 442

RESULT 15
YB04_HAEIN
ID YB04_HAEIN STANDARD; PRT: 407 AA.
AC P71369;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN H1104.
GN H1104.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID: 127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32790; AAC22759.1;
DR TIGR: H1104;
DR InterPro: IPR003662; sub.transprtr.
DR Pfam: PF00083; sugar.tr.1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
```

```
FT TRANSMEM 289 309 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 358 378 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
SQ SEQUENCE 407 AA; 43681 MW; 6185886CB37C4A9 CRC64;

Query Match 9.9%; Score 280.5; DB 1; Length 407;
Best Local Similarity 23.0%; Pred. No. 5e-13;
Matches 107; Conservative 77; Mismatches 185; Indels 97; Gaps 16;

QY 82 FQWKLSTLTGLAMMADAMEMMILSILAPQLHCEWRLPSQVALLTSVVVFGMMSSSTLWG 141
Db 8 YGNKALIGSAVGYGMDGFDLLILGLFMLSISADLNLTPAOGCSLVMTWTLIGAVFGGILFG 67
QY 142 NISI QYGRKTKLSVLWTLTYGILSAFAPVYSWILVLRGLVGFICG--VPQSVTLYAEF 200
Db 68 ALSDKYGRVRVLTWTLILLFAVFTGLCAIQGYWDLIYRTIAGIGLGGEPGIGMALAAEA 127
QY 201 LPMKARAKCILLIEVFWAIGTVFEVLAFAVMPSLGWRWLLILSAVPLILLFAVILCFWLPE 260
Db 128 WPAHRAKAASYVALGHQVGLGAALLTPILLPHIGHRGMLVG-----IFPAFVAVWELR 182
QY 261 SARYDVLSGNOEKAIATFKRIATENGAPMLPGKLIISQEDRGKMRDLFTPHKRWTTLL- 319
Db 183 SHLH-----EPEIFTQKQTA-----LSTQSS-----FTDKLRSFOLL 215
QY 320 -----LWFIWFSNA--FSYVGLVL-----LTTELFAQGDVCGISSRKKAVEAKCSL 363
Db 216 KDKATSKISIGIVVLTQVQNFYGGIWIWLPNLSKQL-----GFSLTQSG----- 261
QY 364 ACEYLSSEEDYMDLLWTLTSEFPQVLTWLTII-----DRLGRKKTALCFVIFSCSLLLI 419
Db 262 -----LWTAVT--VCGMMAGIWFQGLADRIGRKPS-----FLLFOLGAVISIV 303
QY 420 CVGRNVLTLLLTARAFIS-----GGFQAAVYVTEVYPTATRALCLGTCGSMAR--VG 471
Db 304 VYSQLTDPDMLLAGAFGLMFVNGMLGCGYCALMA---EAYPTARATAQNVLFNTGRAVG 360
QY 472 ALITPPTAQVWLESSVYLTAVYSGCCLLAALASCFPIETKGGGL 517
Db 361 GFGPVVVGSVVLAY:FOATALALLAIIVVIDMLATIFILPELKGKAL 406

Search completed: March 13, 2002, 12:42:52
Job time: 158 sec
```

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|--------|-------|-----|--------|--------|--------|--------------|
| | | Match | % | | | | |
| 1 | 2749 | 96.6 | 548 | 11 | Q92217 | Q92217 | rattus norv |
| 2 | 1129.5 | 39.7 | 497 | 5 | Q9W1D4 | Q9W1D4 | drosofila |
| 3 | 863 | 30.3 | 470 | 10 | Q9LSH7 | Q9LSH7 | arabidopsis |
| 4 | 855 | 30.8 | 500 | 10 | Q9NMW2 | Q9NMW2 | arabidopsis |
| 5 | 734 | 25.8 | 144 | 4 | Q9NPW5 | Q9NPW5 | homo sapien |
| 6 | 644.5 | 22.6 | 454 | 2 | Q9RYN9 | Q9RYN9 | deinoceroc |
| 7 | 509.5 | 17.9 | 422 | 2 | Q9X1Y9 | Q9X1Y9 | thermotoga |
| 8 | 466.5 | 16.4 | 455 | 2 | Q9HYD9 | Q9HYD9 | pseudomonas |
| 9 | 446 | 15.7 | 709 | 5 | Q9W3W9 | Q9W3W9 | pseudomonas |
| 10 | 436 | 15.3 | 593 | 11 | P70485 | P70485 | drosofila |
| 11 | 435 | 15.3 | 555 | 11 | Q9ROW2 | Q9ROW2 | rattus norv |
| 12 | 435 | 15.3 | 593 | 11 | P97558 | P97558 | rattus norv |
| 13 | 432 | 15.2 | 535 | 11 | Q63314 | Q63314 | rattus norv |
| 14 | 429.5 | 15.1 | 556 | 11 | Q63089 | Q63089 | rattus norv |
| 15 | 429 | 15.1 | 554 | 6 | O02713 | O02713 | sus scrofa |
| 16 | 427 | 15.0 | 554 | 4 | O15395 | O15395 | homo sapien |
| 17 | 425 | 14.9 | 554 | 4 | O15245 | O15245 | homo sapien |
| 18 | 425 | 14.9 | 556 | 11 | O08966 | O08966 | mus musculus |
| 19 | 423.5 | 14.9 | 556 | 11 | Q9R1Q4 | Q9R1Q4 | mus musculus |

| | | | |
|----|----|--|-----|
| Qy | 1 | MEEDLFQLRQLPVVKFRRTGESARSEDDTASGEHEVQIEGVHVGLBAVELDDGAAVPKEF | 60 |
| Db | 1 | MEEDLFQLRQLPVVKFRRTGESARSEDDAASGEHDVQIEGVHVGLBAVELDDGAAVPKEF | 60 |
| Qy | 61 | ANPTDQTFVYDAVEAIGFGKFWKLVSVLTLGLAWNADAMEMILSLIAPQLHCEWRLPSW | 120 |
| Db | 61 | ANPTDQTFVYDAVEAIGFGRFOWKLVSVLTLGLAWNADAMEMILSLIAPQLHCEWRLPSW | 120 |

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA:
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA:
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 clones.";
 RL DNA Res. 7:131-135(2000).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 DR EMBL; AB026645; BAB02515.1; -;
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 KW Transmembrane.
 SQ SEQUENCE 470 AA; 51457 MW; 1062413BA354616D CRC64;

Query Match 30.3%; Score 863; DB 10; Length 470;
 Best Local Similarity 38.8%; Pred. No. 7-2e-51;
 Matches 180; Conservative 92; Mismatches 168; Indels 24; Gaps 6;
 QY 68 FMVEDAVEAIGFGKFWKLSVLTGLANMADAMEMMILSLAPQLHCEWRLPSPNOVALTS 127
 DB 8 FTVDALVAMGFGKFIQYVLAYAGMGWVAEAMEMMILSFVGPVQSLNLSARQESLTS 67
 QY 128 VVFGMSSSTLWGNISDQYGRKTLKISVLWLTLYGILSAFAPVYSWILVRLGLVGGI 187
 DB 68 VVFGMLIGAYSWGIVSDKHGRKGFITAVVTFVAGFLSAFSPNMMIILRLCLVGLGL 127
 QY 128 VVFGMSSSTLWGNISDQYGRKTLKISVLWLTLYGILSAFAPVYSWILVRLGLVGGI 187
 DB 68 VVFGMLIGAYSWGIVSDKHGRKGFITAVVTFVAGFLSAFSPNMMIILRLCLVGLGL 127
 QY 188 GGVPSQSVTLTAEFLPMKARAKCILLIEVFWAIGTVFVVLAVFVMPSLGWRWLLILSAVP 247
 DB 128 GGGPVLASWYLEFIPAPSRGTWVVFSAFTVGTIFEASLAWLMPRLGWRWLLAFSSVP 187
 QY 248 LLLFVLCFWLPESARDVLSGNOEKATLKRIATENGAPMLGK-----IISROE 300
 DB 188 SLLLLFYRWTSSEPRYLILQGRKAEALALEKIARMKTKQLPGLVLSSELETELENKE 247
 QY 301 DRGMRLDFTPHFRWTTLLWFIWFSNAFSGYGLVLTTELFOAGDVCYGSRRKKAVEAK 360
 DB 248 PGFSLALLSPTLMKRILLWVFFGNFAYYGVVLTTELNNSHNRCYPTEKQ----- 301
 QY 361 CSLACEYLSSEEDYMDLWTTLSLSEPGVVLVLIIDRLGRKTKMALCFVIFSCSLLLFIC 420
 DB 302 ---LRNSNDVNYRDVFIAFSGPGLLSAAMVDRLGRKASMA--SMLETCIFLLPL 355
 QY 421 VGRN--VLTLLLFARAFISGGFOAAVYVPEVPTATRALGLTCSGMARVGALITPFI 478
 DB 356 SHQSPFITVTLFGGRICISAAFTVVIYAYPEIYPTAVRTGTVGSGVSGRIGILCLPV 415
 QY 479 AQVMESSVYLTAV--YSGCCLLAALASCFLPIETKGGLOES 520
 DB 416 A-VGLVHCCHQTIIVLFEVVLVSGICVCLFPFETSGRDLTDS 458

RESULT 4
 Q9M7W2 PRELIMINARY; PRT; 500 AA.
 ID Q9M7W2;
 AC Q9M7W2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PUTATIVE TRANSPORTER.
 GN MGH6.16.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III P1 MGH6 genomic sequence.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 DR EMBL; AC024128; AAF35954.1; -;
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 KW Transmembrane.
 SQ SEQUENCE 500 AA; 54573 MW; EC459EB09EF581C5 CRC64;

Query Match 30.0%; Score 855; DB 10; Length 500;
 Best Local Similarity 36.4%; Pred. No. 2.7e-50;
 Matches 180; Conservative 95; Mismatches 165; Indels 54; Gaps 7;
 QY 68 FMVEDAVEAIGFGKFWKLSVLTGLANMADAMEMMILSLAPQLHCEWRLPSPNOVALTS 127
 DB 8 FTVDALVAMGFGKFIQYVLAYAGMGWVAEAMEMMILSFVGPVQSLNLSARQESLTS 67
 QY 128 VVFGMSSSTLWGNISDQYGRKTLKISVLWLTLYGILSAFAPVYSWILVRLGLVGGI 187
 DB 68 VVFGMLIGAYSWGIVSDKHGRKGFITAVVTFVAGFLSAFSPNMMIILRLCLVGLGL 127
 QY 188 GGVPSQSVTLTAEFLPMKARAKCILLIEVFWAIGTVFVVLAVFVMPSLGWRWLLILSAVP 247
 DB 128 GGGPVLASWYLEFIPAPSRGTWVVFSAFTVGTIFEASLAWLMPRLGWRWLLAFSSVP 187
 QY 248 LLLFVLCFWLPESARDVLSGNOEKATLKRIATENGAPMLG----- 292
 DB 188 SLLLLFYRWTSSEPRYLILQGRKAEALALEKIARMKTKQLPGLVLSSELETELENKE 247
 QY 293 -----KLIISROEDRG-KMRDLFTPHFRWTTLLWFIWFSNAF 330
 DB 248 IPTENTHLLKAGESGEAVAVSKIVLKADKEPGFSLALLSPTLMKRILLWVFFGNF 307
 QY 331 YGVLVLTTELFOAGDVCYGSRRKKAVEAKCSLACEYLSSEEDYMDLWTTLSLSEPGVLT 390
 DB 308 YGVVLTTELNNSHNRCYPTEKQ-----LRNSNDVNYRDVFIAFSGPGLLS 357
 QY 391 LWIIDRLGRKTKMALCFVIFSCSLLLFICVGRN--VLTLLLFARAFISGGFOAAVYV 448
 DB 358 AAMVDRLGRKASMA--SMLETCIFLLPLLSHQSPFITVTLFGGRICISAAFTVVIYA 415
 QY 449 PEVPTATRALGLTCSGMARVGALITPFIQVMESSVYLTAV--YSGCCLLAALASC 506
 DB 416 PEIYPTAVRTGTVGSGVSGRIGILCLPLA-VGLVHCCHQTIIVLFEVVLVSGICVC 474
 QY 507 FLPIETKGGLOES 520
 DB 475 LFPFETSGRDLTDS 488

RESULT 5
 Q9NPW5 PRELIMINARY; PRT; 144 AA.
 ID Q9NPW5;
 AC Q9NPW5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 15.4 KDA PROTEIN (FRAGMENT).
 GN DKE2P761H039.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=AMYGDLA;
 RL Blum H., Bauersachs S., Mewes H.W., Weill B., Wiemann S.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL359592; CAB94878.1; -
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 144 AA; 15387 MW; 769A51A9564FFCD CRC64;
 Query Match 25.8%; Score 734; DB 4; Length 144;
 Best Local Similarity 99.3%; Pred. No. 1e-42;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 405 LCFVIFSCSLLLCVGRNVLTLLFTARAFISGFGQAAAYVYTPVPTATRALGLGTC 464
 DB 1 LCFVIFSCSLLLCVGRNVLTLLFTARAFISGFGQAAAYVYTPVPTATRALGLGTC 60
 OY 465 SGMARVALITPPIAQVMLESSVYTLAVYSGCCLLAALASCFLPIETKGGGLOESSHRE 524
 DB 61 SGMARVALITPPIAQVMLESSVYTLAVYSGCCLLAALASCFLPIETKGGGLOESSHRE 120
 OY 525 WQGMVGRGMHGAGVTRNSGSQ 548
 DB 121 WQGMVGRGMHGAGVTRNSGSQ 144
 RESULT 6
 OQRYN9 PRELIMINARY; PRT; 454 AA.
 AC OQRYN9;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE SUGAR TRANSPORTER, PUTATIVE.
 GN DRA0271.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 DR EMBL; AE001863; AAF12486.1; -
 DR TIGR; DRA0271; -
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar.tr. 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 KW Complete proteome; Sugar transporter; Transmembrane.
 SQ SEQUENCE 454 AA; 48171 MW; 214EA1A3EDC60B8B CRC64;
 Query Match 22.6%; Score 644.5; DB 2; Length 454;
 Best Local Similarity 30.6%; Pred. No. 4.5e-36;
 Matches 142; Conservative 95; Mismatches 190; Indels 37; Gaps 5;

OY 70 VEDAVEAIGFGKFWKLSVLTGLAWNADAMEMMILSILAPQLHCEWRPLPSQOVA--LLTS 127
 DB 15 VDRVUDDLGLGRFQWKLAIICGLTWAADAMEVLLMGFALPGISAAPFELPKGSPAATMLLT 74
 OY 128 VVFVGMSSSTLWGNISDOYGRKTKGLKISVLWTLVYTGILSAPAPVYSWILVRLGLVGF 187
 DB 75 ATFAGMLFGANFWGYLADRVGRRSVFLTTVALGVVFGLAGALAPTLTLLVARFLTGFAI 134
 OY 198 GG-VPOSVTLVLAELPMKARAKCILLIEVFAIGVVEVVLAVFV---MPSLGRWRLLI 242
 DB 135 GGTLPVDYSMMAEFPTAWGRFLVYLFESFVWVAAALAWWYSTAFAPAEGRWLLG 194
 OY 243 LSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIATLKRIATENGAMPILKGLIISQEDR 302
 DB 195 LAALPGLVGLIARIIGIPDSPRSLLARGEAAQARAALQKVAANGGTLTAPLAHPQPPR 254
 OY 303 GKMDLFTPHRPWTLLWFWFNSNAFSYGLV-----LTTTFLFQAGDVCGISSRKNV 357
 DB 255 VSPAQLFRGVLAARRTPLLMVTWFGSLSGYIGFISWLPFSFLRAQGLDLGAV----- 304
 OY 358 EAKCSLACEYLSEEDYMDLLTTLTTLSEFPGLVLTWIIDRLGRKKTMAICFVIFSCSLL 417
 DB 305 -----YRSTLLALLAQVPCYLLAAVLEKIGRRVTLVGLTLGAVGAYLF 349
 OY 418 FICVGRNVLTLLFTARAFISGFGQAAAYVYTPVPTATRALGLGTCSCMARVAGALITPF 477
 DB 350 LLAHDANTVLTLSALLFALLGANGSLAYTPELPTPLRTTGMGLVSGVARLASVSPS 409
 OY 478 IAOVMLESSVYTLAVYSGCCLLAALASCFLPIETKGGGLOESS 521
 DB 410 IGAMLLTGNLTALIVFAVCFALAAALAAWIGVETRGQALAE 453
 RESULT 7
 OQX1T9 PRELIMINARY; PRT; 422 AA.
 AC OQX1T9;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE PERMEASE, PUTATIVE.
 GN TM1603.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
 DR EMBL; AE001804; AAD36670.1; -
 DR TIGR; TM1603; -
 DR InterPro; IPR000515; BPD_transp.
 DR Pfam; PF00083; sugar.tr. 1.
 DR PROSITE; PS00402; BPD_TRANSPORT_1; UNKNOWN_1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
 KW Complete proteome; Transmembrane.
 SQ SEQUENCE 422 AA; 47152 MW; 13F9DC5649A1338D CRC64;
 Query Match 17.9%; Score 509.5; DB 2; Length 422;
 Best Local Similarity 29.5%; Pred. No. 5.9e-27;

| | | |
|----------|---|---|
| | | Matches 135; Conservative 81; Mismatches 188; Indels 53; Gaps 11; |
| QY | 70 | VEDAVEAIGFGKFWKLSVLTGLAWADAMEMMILSILAPOLHCEWRKLPSSQWQVALLTSVV 129 |
| Db | 3 | IDIVEIKYVDRKTORRFLITSIAMFDDAGVWLLSFVLPYVVIKENWLTSTOGATIASAT 62 |
| QY | 130 | FVGMSSSTLWGNISDQYGRKTGLKIS-VLWTLTYGILSAFAPYVSWILVRLGLVCGIGG 189 |
| Db | 63 | FLCMLFALSFGVADLLGRKVSNNLFFVITFTFLSGFSSSFETLLVRLGLSGFGYGG 122 |
| QY | 190 | VPOSVTLV-AEFLPMKARAKCILLIEVFWAIGTVFVVLAVFVMPSLGWRWLLLSAVPL 248 |
| Db | 123 | IMPSEFNAYLAETSIIRGRYLVLLSSWAGVSLILGLFVAVNLPN--WRWFWIFSIG- 179 |
| QY | 249 | LLFAVLFCWLPESARYDVLSSGNOEKAIATL-KRIATENGAP-----MPLGKLIIISROEDRG 303 |
| Db | 180 | YLFVVPFLRMPETPKYAFLLKGGKEALERSLGRVEEVEVLPPKKEKVPILALL-----KRE 234 |
| QY | 304 | KMRDLFTPHRWTLTLWFLWFSNATSYGLVLLTTLTFOAGDVCGISSRKKAVEAKCSL 363 |
| Db | 235 | HLKD-----TVVIWIAWFVSEVYVYALFTWAPRISS-----LGVSVYKSS----- 275 |
| QY | 364 | ACEVLSSEYDMLWTT-----LSEFFPGLVTLWIIDRLGRKKTMAFCVIFSPCSLLLFI 419 |
| Db | 276 | -----WTFYMWVAQLPGYLSAAFYIEKWRKKSGLSYVFTGTLAALLWAN 321 |
| QY | 420 | CVGRNVLTLTLFIARAFISGGFOAAVYVTPETVATRALGLTCSGMARVALITPPTIA 479 |
| Db | 322 | VRGDASLLAAALVSLPFCGLVWGLVAYTPELYPTSLRGTTGGAAGVWARIAGIITAPYTT 381 |
| QY | 480 | QVMLE---SSVYLVTLAVYSGCLLAALASCFLEPIETKG 514 |
| Db | 382 | GFMMEXGKSAETLAWISAMAFAGVILVIFGRETGK 418 |
| RESULT 8 | | |
| Q9HYD9 | PRELIMINARY; PRT; 455 AA. | |
| AC | Q9HYD9; 16, Created | |
| DT | 01-MAR-2001 (Tremblrel. 16, Last sequence update) | |
| DT | 01-MAR-2001 (Tremblrel. 16, Last sequence update) | |
| DE | PROBABLE MFS TRANSPORTER. | |
| GN | PA3467. | |
| OS | Pseudomonas aeruginosa | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; | |
| OC | Pseudomonas | |
| OX | NCBI_TaxID=287; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=PA01; | |
| RX | MEDLINE=20437337; PubMed=10984043; | |
| RA | Stover C.K., Pham Q.-T., Erwin A.L., Mizoguchi S.D., Warrenner P., | |
| RA | Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., | |
| RA | Garber R.L., Gentry L., Tolentino E., Westbrook-Wadman S., Yuan Y., | |
| RA | Brody L.L., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M., | |
| RA | Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T., | |
| RA | Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.; | |
| RA | "Complete genome sequence of Pseudomonas aeruginosa PA01, an | |
| RA | opportunistic pathogen." | |
| RT | Nature 406:959-964(2000). | |
| RL | -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY). | |
| CC | -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY. | |
| DR | ENBL; AE004767; AAG06855.1; - | |
| DR | InterPro: IPR003662; sub.transporter. | |
| DR | Pfam: PF00083; sugar.tr.1 | |
| DR | PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1. | |
| DR | PROSITE; PS00217; SUGAR_TRANSPORT_2; 1. | |
| KW | Complete proteome; Transmembrane. | |
| SQ | SEQUENCE 455 AA; 49357 MW; 2727ED655908A400 CRC64; | |

| | | |
|----------|--|---|
| | | Matches 132; Conservative 94; Mismatches 189; Indels 61; Gaps 13; |
| QY | 72 | DAVEAIGFGKFWKLSVLTGLAWADAMEMMILSILAPOLHCEWRKLPSSQWQVALLTSVV 131 |
| Db | 10 | ERLERLPLSPYHRLVPIIALAFFDSDMLAMWTFLLGSIKAEFGDLSAQAGLLASSSFF 69 |
| QY | 132 | GMSSSTLWGNISDQYGRKTGLKIS-VLWTLTYGILSAFAPYVSWILVRLGLVCGIGG-G 189 |
| Db | 70 | GMVIGALSGLMLADRFGRKPVFQASIVLWGLASYLCSTAGDLDLS-LTFYRVLLIGIGMGE 128 |
| QY | 190 | VPOSVTLV-AEFLPMKARAKCILLIEVFWAIGTVFVVLAVFVMPSLGWRWLLLSAVPL 249 |
| Db | 129 | FPIAQSLLEMPASRGKIALMDGFNPLGFVAACLSYFLPLTGWRSIFLVLPALPAV 188 |
| QY | 250 | LFVAVLCWLPESARYDVLSSGNOEKAIATLKRIATE-----NCAPMPICKLIIISQEDR 302 |
| Db | 189 | FVLAIRFLIPESPRWLEQAGRRQADRLVLRDIARVWRSIGLTELPPPL-----ROPQR 242 |
| QY | 303 | GKMR-----DLFTPHRWTLTLWFLWFSNATSYGLVLLTTLTFOAGDVCGISSRK 354 |
| Db | 243 | ERSRPGFFSAFELWSPAYRRRTLTVMGLFFALLGFGYGLTWSLSALLQSGFA----- 296 |
| QY | 355 | KAVEAKCSLACEVLSSEYDMLWTTLSLSEFFPGLVTLWIIDRLGRKKTMAFCVIFSPCSLLLFI 409 |
| Db | 297 | -----VTOSVYVTVL- ISLAGIPGLCAAWLVESWGRKPSCVLMLLGGGAMA 342 |
| QY | 410 | FSCSLLLFICVGRNVLTLTLFIARAFISGGFOAAVYVTPETVATRALGLTCSGM 467 |
| Db | 343 | YAYGQTAVF---GGSLLALIGFLAMQFFLFGMAVLYTYTPELYPTSLRGTTGGAAGVWARIAGIITAPYTT 399 |
| QY | 468 | ARVAGLITPPIAOVML---ESSVYLVTLAVYSGCLLAALASCFLEPIETKG 519 |
| Db | 400 | GRIGSLGLPLVTGLVPLTGCQGVFTLGLAL---CFGVAALVWVAFGIETRGRTLEE 452 |
| RESULT 9 | | |
| Q9W3W9 | PRELIMINARY; PRT; 709 AA. | |
| AC | Q9W3W9; 13, Created | |
| DT | 01-MAY-2000 (Tremblrel. 13, Created) | |
| DT | 01-MAY-2000 (Tremblrel. 13, Last sequence update) | |
| DE | CG3168 PROTEIN. | |
| GN | CG3168. | |
| OS | Drosophila melanogaster (Fruit fly). | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | |
| OC | Pterygota; Neoptera; Diptera; Brachycera; Muscomorpha; | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | |
| OX | NCBI_TaxID=7227; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=BERKELEY. | |
| RX | MEDLINE=20196006; PubMed=10731132; | |
| RA | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., | |
| RA | Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., | |
| RA | George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N., | |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | |
| RA | Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., | |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., | |
| RA | Aquil J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D., | |
| RA | Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | |
| RA | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | |
| RA | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | |
| RA | Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., | |
| RA | Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., | |
| RA | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K., | |
| RA | Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., | |
| RA | Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., | |

| | | |
|-------------|--|---------------------------------------|
| Query Match | | 16.4%; Score 466.5; DB 2; Length 455; |
|-------------|--|---------------------------------------|

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rebert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Rhue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: AE003438; AAF46193.1; -.
DR FlyBase: FBgn0029896; CG3168.
DR InterPro: IPR003662; sub.transprotector.
DR Pfam: PF00083; sugar tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane.
SQ SEQUENCE 709 AA; 77930 MW; 1B5AD1E9D133AE94 CRC64;

Query Match 15.7%; Score 446; DB 5; Length 709;
Best Local Similarity 25.5%; Pred. No. 2.2e-22;
Matches 127; Conservative 98; Mismatches 210; Indels 64; Gaps 12;
QY 71 EDVAEAFGKQKQKLVLTGLAWMADAMEMMILSILAPQLHCEWRLPSQVALLTSVVF 130
DB 215 ERAIELCGYKPHYILLAIAGLSTSEEMDVMSFILPSAECDLDTETKGLNSIIF 274
QY 131 VGMSSSTLGNISDOYGRKTKLISVLTLYGILSAFAPYNSILVRLGLVGGIGV 190
DB 275 IGMWYGVFWGSIADSGRKKVLIIVISFMNFCIVASSFSQTSYFFMLFRFLNGAALGS 334
QY 191 -PQSVTLAEFLPMKARAKCILLIEVFWAIGTVFVAVFVMP-SLG-----WR 238
DB 335 GPVWSYFAEFQPKAKRGSMLSFMAAFWTFGNLFVASLAWLIIPRTGTFPTFYNSWR 394
QY 239 WLLILSAVPLLLFAVLCTWLPESARYDVLGNGQEKATLKRIATENGAPMLGLIISR 298
DB 395 IFLLVCSLPFLVGLFLLFYPSPKFLTRGKKDRALAFRGIFVTNKKRPDEYMYVDL 454
QY 299 Q-----EDRGKMRDLF-----TPHFRWTTLLLWFIFSNFAYVGLVL 336
DB 455 EYDEKLLSNGVNNKYSRMISGMVDHSLRALKSPILRFTIVSI-TINFTHIGYGLLM 513
QY 337 LTTELF-----OAGDVCGIS-----SRKKAKEAKCSLACEYLSSEEDYMDLLW 378
DB 514 WPELFNRFEEYEKAFPPQSGAGCAVTVVYNLAKEQSNNGTCS---SDIQSVEMESLI 570
QY 379 TTLSEFPGLVTLWILDRGRKKTALCFVIFSCSLILFICVGRNVLTLLFIARAFIS 438
DB 571 SLASALPAMLAIGDMGLGRKFFLIAGTMTAGICSALMYF-VRSSVONL-----VVSATFS 626
QY 439 GGFQAAVY-----YTPVYPTATRALGLGTCSGMARVGALITPFIQVMSLESSVYLTAVY 494
DB 627 GAISAANAALDCLITEVPTKLRTGVALISMVAARLGGIIGNIVIAQLLDNVCPSPTFIV 686
QY 495 SCCCLLAALASCFIDIEK 513
DB 687 SGLLGGGLMCLLLPNTTR 705
RESULT 10

P70485
ID P70485; PRELIMINARY; PRT: 593 AA.
AC Q9ROW2; PRELIMINARY; PRT: 555 AA.
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update).
DE ORGANIC CATION TRANSPORTER OCT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RX MEDLINE=96295517; PubMed=8702418;
RT Okuda M., Saio H., Urakami Y., Takano M., Inui K.;
RT "cDNA cloning and functional expression of a novel rat kidney organic
RT cation transporter, OCT2."
RL Biochem. Biophys. Res. Commun. 224:500-507(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: D83044; BAAL1754.1; -.
DR InterPro: IPR003662; sub.transprotector.
DR Pfam: PF00083; sugar tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Transmembrane.
SQ SEQUENCE 593 AA; 66080 MW; 36C1044E0C04B443 CRC64;
Query Match 15.3%; Score 436; DB 11; Length 593;
Best Local Similarity 29.5%; Pred. No. 8.4e-22;
Matches 127; Conservative 77; Mismatches 148; Indels 78; Gaps 16;
QY 119 SNOVAITLTVFVGMSSSTLGNISDOYGRKTKLISVLTLYGILSAFAPYNSILV 178
DB 146 SMLDLFQSVWVNGVFFIGAMMIGYLADRGKRCCLLVTLINAISSALMAISPNAWMLV 205
QY 179 ---LRGLV---GFGIGGVPSVTLVYAEFLPMKARAKCILLIEVFWAIGTVFVVLVLA--VF 230
DB 206 FRFLOGLYSKAGWLLGYI-----LITEFVGLGRVWVGICYQIAFTVGL--LILAGVAY 257
QY 231 VMPISLGRWMLLILSAVPLLLFAVLCTWLPESARYDVLGNGQEKATLKRIATENGAPM 289
DB 258 VIPN--WRWLPQAVTLPNFCF--LLYFWCIPESPRWLLISQNKIVKAMKIIKHAKNGKSV 314
QY 290 PLGKLIIISQEDRGK-----MRDLF--TPHFRWTTLLLWFIFSNFAYVGLVLTTELFQ 343
DB 315 PVSQNLTLDPEDAGKRLKPSIDLVRTPOIRKHTLLIMYNWFTSSVLYOGLIM---HMGL 371
QY 344 AGDVCGIISRRKKKAVEAKCSLACEYLSSEEDYMDLLWTTLSEFFGVLTWILDRGRKKTW 403
DB 372 AGD-----NIYLDFFYSALVEFFAAFIILITIDRVGRYPW 407
QY 404 ALC-FVIFSCSLILFI-----CVCGRNVLTLLLFIARAFISGFGQAAVYVTP 450
DB 408 AVSNVAGAACLASVFIIPDDQLQMLKTIACLRMGITM-----AYEMVCLVNAE 456
QY 451 VYPTATRALGLGTCSGMARVGALITPFIQVMSLESSVYLTAVYSGCCLLAALASCFIPI 510
DB 457 LYPTIYRNLGVLCSSMCDIGIITPFLVYRLTDIMPEPLVVFVAVGLVAGALLLP- 515
QY 511 ETKGGGLQES 520
DB 516 ETKGKALPET 525
RESULT 11
QY 516 ETKGKALPET 525
ID Q9ROW2; PRELIMINARY; PRT: 555 AA.
AC Q9ROW2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE ORGANIC CATION TRANSPORTER OCT2R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=99316020; PubMed=10385678;
RA Gruendemann D., Liebig G., Kiefer N., Koester S., Schoenig E.;
RT "Selective substrates for non-neuronal monoamine transporters.";
RL Mol. Pharmacol. 56:1-10(1999)
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; Y13154; CAB52215.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR Transmembrane.
SQ SEQUENCE 555 AA; 62342 MW; 29521969AEIAC206 CRC64;

Query Match 15.3%; Score 435; DB 11; Length 555;
Best Local Similarity 29.5%; Pred. No. 9.1e-22;
Matches 127; Conservative 76; Mismatches 149; Indels 78; Gaps 16;

QY 119 SNOVALLTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGLISAFAPVYSMLV 178
DB 119 SNOVALLTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGLISAFAPVYSMLV 178
DB 146 SMLDLFQSVVNVGFFIGAMMIGYLADRFGRKFCLLVTLINAISSGALMAISPNYAMLV 205
QY 179 ---LRGLV---GFGIGGVPOSVTLVYAEFLPMKARAKCILLIEVFAIGTVFEVLA--VF 230
DB 179 ---LRGLV---GFGIGGVPOSVTLVYAEFLPMKARAKCILLIEVFAIGTVFEVLA--VF 230
DB 206 FRFQGLVSKAGWLGVI-----LITEFVLGYRRVMVGICYQIAFTVGL---LILAGVAY 257
QY 231 VPSLGRWMLLILSAVPLLLFAVLFCW-LPESARYDVLGSGNQEKATATLKRIATENGAPM 289
DB 231 VPSLGRWMLLILSAVPLLLFAVLFCW-LPESARYDVLGSGNQEKATATLKRIATENGAPM 289
DB 258 VIPN--WRMLQFAVTLPNFCF--LLYFWCIPESPRWLISQNKIVKAMKIIKHAKKNGKSV 314
QY 290 PLGKLIISROEDRGK-----MRDLF--TPHFRWTTLLLWFIWFSNAFYSGVLVLTLELQ 343
DB 290 PLGKLIISROEDRGK-----MRDLF--TPHFRWTTLLLWFIWFSNAFYSGVLVLTLELQ 343
DB 315 PVSQNLNTPDEDAGKLNPSFLDVRTPQIRKHTLILMYNWFSSVLYOGLIM---HMGL 371
QY 344 AGDVCCISSRKKAVEAKSLACEYLSEEDYMDLLWTLTSEFPGLVLTWIIDRLGRKKT 403
DB 344 AGDVCCISSRKKAVEAKSLACEYLSEEDYMDLLWTLTSEFPGLVLTWIIDRLGRKKT 403
DB 372 AGD-----NIYDFYSALVEFPAAFIILITIDRVGRYPW 407
QY 404 ALC-FVIFSFCSLLFI-----CVGRNVLTLLFIARAFISGFGQAAYVYTP 450
DB 404 ALC-FVIFSFCSLLFI-----CVGRNVLTLLFIARAFISGFGQAAYVYTP 450
DB 408 AVSNVAGAACLASVFIPLDQWLKTIACLRMGITM-----AYEMVCLVNAE 456
QY 451 VYPATRALGLGTCGSMARVGLITPFIAQVLMLESSVYLTAYISGCLLAALASCFLPI 510
DB 451 VYPATRALGLGTCGSMARVGLITPFIAQVLMLESSVYLTAYISGCLLAALASCFLPI 510
DB 457 LYPTIIRNLGLVSCSSMCDIGGIITPFLVYRLTDIWMFPLVVFVAVGLVAGALVLLLP- 515
QY 511 ETKGGGLOES 520
DB 511 ETKGGGLOES 520
DB 516 ETKGKALPET 525

RESULT 12
P97558 PRELIMINARY; PRT; 593 AA.
ID P97558
AC P97558;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ORGANIC CATION TRANSPORTER.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=KIDNEY;
RA Gorboulev V.G., Koepsell H.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; X98334; CAA66979.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR Ion transport; Transmembrane.
SQ SEQUENCE 593 AA; 66100 MW; 36C3E1B5DC057790 CRC64;

Query Match 15.3%; Score 435; DB 11; Length 593;
Best Local Similarity 29.5%; Pred. No. 9.8e-22;
Matches 127; Conservative 76; Mismatches 149; Indels 78; Gaps 16;

QY 119 SNOVALLTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGLISAFAPVYSMLV 178
DB 119 SNOVALLTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGLISAFAPVYSMLV 178
DB 146 SMLDLFQSVVNVGFFIGAMMIGYLADRFGRKFCLLVTLINAISSGALMAISPNYAMLV 205
QY 179 ---LRGLV---GFGIGGVPOSVTLVYAEFLPMKARAKCILLIEVFAIGTVFEVLA--VF 230
DB 179 ---LRGLV---GFGIGGVPOSVTLVYAEFLPMKARAKCILLIEVFAIGTVFEVLA--VF 230
DB 206 FRFQGLVSKAGWLGVI-----LITEFVLGYRRVMVGICYQIAFTVGL---LILAGVAY 257
QY 231 VPSLGRWMLLILSAVPLLLFAVLFCW-LPESARYDVLGSGNQEKATATLKRIATENGAPM 289
DB 231 VPSLGRWMLLILSAVPLLLFAVLFCW-LPESARYDVLGSGNQEKATATLKRIATENGAPM 289
DB 258 VIPN--WRMLQFAVTLPNFCF--LLYFWCIPESPRWLISQNKIVKAMKIIKHAKKNGKSV 314
QY 290 PLGKLIISROEDRGK-----MRDLF--TPHFRWTTLLLWFIWFSNAFYSGVLVLTLELQ 343
DB 290 PLGKLIISROEDRGK-----MRDLF--TPHFRWTTLLLWFIWFSNAFYSGVLVLTLELQ 343
DB 315 PVSQNLNTPDEDAGKLNPSFLDVRTPQIRKHTLILMYNWFSSVLYOGLIM---HMGL 371
QY 344 AGDVCCISSRKKAVEAKSLACEYLSEEDYMDLLWTLTSEFPGLVLTWIIDRLGRKKT 403
DB 344 AGDVCCISSRKKAVEAKSLACEYLSEEDYMDLLWTLTSEFPGLVLTWIIDRLGRKKT 403
DB 372 AGD-----NIYDFYSALVEFPAAFIILITIDRVGRYPW 407
QY 404 ALC-FVIFSFCSLLFI-----CVGRNVLTLLFIARAFISGFGQAAYVYTP 450
DB 404 ALC-FVIFSFCSLLFI-----CVGRNVLTLLFIARAFISGFGQAAYVYTP 450
DB 408 AVSNVAGAACLASVFIPLDQWLKTIACLRMGITM-----AYEMVCLVNAE 456
QY 451 VYPATRALGLGTCGSMARVGLITPFIAQVLMLESSVYLTAYISGCLLAALASCFLPI 510
DB 451 VYPATRALGLGTCGSMARVGLITPFIAQVLMLESSVYLTAYISGCLLAALASCFLPI 510
DB 457 LYPTIIRNLGLVSCSSMCDIGGIITPFLVYRLTDIWMFPLVVFVAVGLVAGALVLLLP- 515
QY 511 ETKGGGLOES 520
DB 511 ETKGGGLOES 520
DB 516 ETKGKALPET 525

RESULT 13
Q63314 PRELIMINARY; PRT; 535 AA.
ID Q63314
AC Q63314;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LIVER-SPECIFIC TRANSPORT PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE=94334371; PubMed=8056831;
RA Simonson G.D., Vincent A.C., Roberg K.J., Huang Y., Iwanij V.;
RT "Molecular cloning and characterization of a novel liver-specific transport protein.";
RT J. Cell Sci. 107:1065-1072(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; L27651; AAA57157.1; -.

| | |
|---|---|
| DR | InterPro: IPR003662; sub_transporter. |
| DR | Pfam: PF00083; sugar_tr; 1. |
| DR | Transmembrane. |
| KW | SEQUENCE 535 AA; 58712 MW; ESC1896BDEAE2B73 CRC64; |
| SQ | SEQUENCE 535 AA; 58712 MW; ESC1896BDEAE2B73 CRC64; |
| | |
| Query Match | 15.2%; Score 432; DB 11; Length 535; |
| Best Local Similarity | 28.7%; Pred. No. 1.4e-21; |
| Matches 149; Conservative 79; Mismatches 215; Indels 76; Gaps 17; | |
| QY | 34 HEVQISGVHVGLEAVELDDCAAPVPREFANPTDFTF-----VEDAVEAIGFG 80 |
| DB | 1 HHCALPCAPANLSHODLVAEHLPRE-----TDCSFSSCLRFAYPQTVPNWTGLTEVSNSG 102 |
| QY | 81 KFOWK-LSVLTGLAWMADAMENMILSILAPOLCHEWRPLSWOVAL--LTSV-VFVGMMS 136 |
| DB | 103 EPGEPLTYPCSGOWEYDRSE-----FSSTIATEMDLVCCQORGLNKITSTCFFIGVLVG 156 |
| QY | 137 STLWGNISDYGRKTKLGKISVLWTLYYGILSAFAFYYSMLVRLGLVGFGGVGPQSV-T 195 |
| DB | 157 AVVGYLSDFRGRRLLLVAVVSLVGLMSASINYPIMPVTITLGSALAGFTIIIVLP 216 |
| QY | 196 LYAEFLPMKARAKCIILLIEVFWAIGTVFEVLAVFVMPSSLGWRLTLISAVPLLFLFAVLC 255 |
| DB | 217 LEWLDEVRHTVAGVISTVFNWS-GVGLLLALVGLIRS-WRWLLAAATLPCVPGIISI 273 |
| QY | 256 FWLPESARYDLVSGNOKEATIKRIATENGAMPGLGKL-----IISRQEDRGKMR 306 |
| DB | 274 WWVPESARWLLTGQRVEBEAKKYLLSCAKNGRPVGEGLSQEAALNNVNMTERALQPSYL 333 |
| QY | 307 DLF-TPHFRWTTLLWFIFSNAFSYGLVLLTTELFOAGDVCGISSRKKAWEAKCSLAC 365 |
| DB | 334 DLFRTSQLRHISLCMMWFGVNFYSYGLTL-----DVSG----- 368 |
| QY | 366 EYLSEEDYMOLLTWTLSEFPGLVTLWIIDRLGRKKTM-----LCFVIPFCSLLLFCVGR 423 |
| DB | 369 --LGLNVYTQLLPGAVELPSKIMVFLVRLGRRLTEAGMLLGAALTFTGSLVLSLETK 426 |
| QY | 424 NVTLTLFIARAFISGFOAQAYVTPVYATRALGLTCGSMARVGLITPFIAQVML 483 |
| DB | 427 SWITALVVVGKAFSEAFTTAIFYLSELYPTVLROTGLGTALMGRLCASLARLAA---L 483 |
| QY | 484 ESSVYLT--AVYSGCCLLAALASCFLPIETKGGLQES 520 |
| DB | 484 LDGVWLLPKVAYGGIALVAECTALLP-ETKRQAQLPET 521 |
| | |
| RESULT 14 | |
| ID Q63089 | PRELIMINARY; PRT; 556 AA. |
| AC Q63089; | |
| DT 01-NOV-1996 | (TrEMBLrel. 01, Created) |
| DT 01-NOV-1996 | (TrEMBLrel. 01, Last sequence update) |
| DT 01-JUN-2001 | (TrEMBLrel. 17, Last annotation update) |
| DE ORGANIC CATION TRANSPORTER. | |
| OS Rattus norvegicus (Rat). | |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | |
| OX NCBI_TaxID=10116; | |
| RN [1] | Nature 372:549-552(1994). |
| RP SEQUENCE FROM N.A. | |
| RC TISSUE=LIVER. | |
| RX MEDLINE=95082907; PubMed=7990927; | |
| RA Gruendemann D., Gorboulev V., Ganbarian S., Koepsell H.; | |
| RT "Drug excretion mediated by a new prototype of polyspecific transporter."; | |
| RL Nature 372:549-552(1994). | |
| RN [2] | |
| RP SEQUENCE FROM N.A. | |
| RC TISSUE=LIVER. | |
| RA Gruendemann D., Gorboulev V.; | |
| RA Gorboulev V.G.; | |
| RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases. | |
| RN [3] | |

Wed Mar 13 13:08:17 2002

CC
CC
CC
CC

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
-1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.

CC EMBL; Y09400; CAA70567.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR Transmembrane.
KW SEQUENCE 554 AA; 61989 MW; E14B5565600C553B CRC64;

Query Match 15.1%; Score 429; DB 6; Length 554;
Best Local Similarity 30.3%; Pred. NO. 2.3e-21;
Matches 130; Conservative 65; Mismatches 158; Indels 76; Gaps 14;

[illegible]

Search completed: March 13, 2002, 12:42:29
Job time: 135 sec